# Essays on the Economics of Medical Practice Variations

Ingemar Eckerlund

AKADEMISK AVHANDLING

som för avläggande av filosofie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
fredagen den 16 november 2001, kl 09.15
i sal Ragnar, Handelshögskolan
Sveavägen 65

# Essays on the Economics of Medical Practice Variations

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

***EFI Mission***

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.

EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

***Research Organization***

The research activities are organized in twenty Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

| | |
|---|---|
| *ORGANIZATION AND MANAGEMENT* | |
| Management and Organisation; (A) | Prof Sven-Erik Sjöstrand |
| Center for Ethics and Economics; (CEE) | Adj Prof Hans de Geer |
| Center for Entrepreneurship and Business Creation; (E) | Prof Carin Holmquist |
| Public Management; (F) | Prof Nils Brunsson |
| Information Management; (I) | Prof Mats Lundeberg |
| Center for People and Organization; (PMO) | Prof Jan Löwstedt |
| Center for Innovation and Operations Management; (T) | Prof Christer Karlsson |
| *ECONOMIC PSYCHOLOGY* | |
| Center for Risk Research; (CFR) | Prof Lennart Sjöberg |
| Economic Psychology; (P) | Prof Lennart Sjöberg |
| *MARKETING* | |
| Center for Information and Communication Research; (CIC) | Adj Prof Bertil Thorngren |
| Center for Consumer Marketing; (CCM) | Acting Prof Magnus Söderlund |
| Marketing, Distribution and Industrial Dynamics; (D) | Prof Lars-Gunnar Mattsson |
| *ACCOUNTING, CONTROL AND CORPORATE FINANCE* | |
| Accounting and Managerial Finance; (B) | Prof Lars Östman |
| Managerial Economics; (C) | Prof Peter Jennergren |
| *FINANCE* | |
| Finance; (FI) | Prof Clas Bergström |
| *ECONOMICS* | |
| Center for Health Economics; (CHE) | Prof Bengt Jönsson |
| International Economics and Geography; (IEG) | Prof Mats Lundahl |
| Economics; (S) | Prof Lars Bergman |
| *ECONOMICS STATISTICS* | |
| Economic Statistics; (ES) | Prof Anders Westlund |
| *LAW* | |
| Law; (RV) | Prof Erik Nerep |

*Chairman of the Board:* Prof Håkan Lindgren. *Director:* Associate Prof Bo Sellstedt.

***Adress***

EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

# Essays on the Economics of Medical Practice Variations

Ingemar Eckerlund

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics, 2001.


ISBN nr 91-7258-583-8

*Keywords:*
Cesarean section, medical practice variations, economics, regression analysis, health outcome, economic consequences, efficiency, productivity, data envelopment analysis, patient satisfaction, cost-effectiveness, cost-benefit, willingness-to-pay.

*Printed by:*
Elanders Gotab, Stockholm 2001.

*Distributed by:*
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se/efi

# Acknowledgements

During my work on this thesis, I have benefited from the help of a number of very competent and nice people. First of all, I want to thank my supervisor, Professor Bengt Jönsson, who always generously shared with me his richness in knowledge and experience, and often reminded me of the importance of a good economic theory. Without his support and encouragement, this volume would probably not have been written.

I am also much indebted to my statistical adviser and co-author, Ulf Gerdtham, for guiding me through the jungle of statistical test methods. Likewise, I am very grateful to other colleagues at the Centre for Health Economics, for nice company and for many helpful comments on early drafts of my papers: Magnus Tambour, co-author of one of the papers, Magnus Johannesson, Per-Olov Johansson, Claes Rehnberg, Niklas Zethraeus, Göran Karlsson and Carin Blanksvärd.

Thanks to Anders Westlund and Jan Eklöf at the Department of Economic Statistics, I became involved in some very stimulating projects on consumer (patient) satisfaction. The result was two of the papers in this thesis. Claes Cassel explained the partial least squares technique to me.

I also want to express my gratitude to Egon Jonsson and Stefan Håkansson, who share the responsibility for first directing me to the subject of medical practice variation.

A prerequisite for applied health economics is a close collaboration with the medical profession. Stig Gårdmark first guided me into obstetrics and gynecology, and Jörgen Nathorst-Böös, co-author of the last paper, took me further in that exiting field. My interest in ophthalmology is a consequence of many years of stimulating collaboration with Klas Göran Brege and William Thorburn.

Financial support from the National Corporation of Swedish Pharmacies, the Swedish Council for Social Research, The Swedish Institute for Health Services Development and the National Board of Health and Welfare is gratefully acknowledged.

Finally, my greatest thanks to my wife, Kristin, and our children, Sara, Moa and Kalle, for their loyalty, support and patience.

Stockholm, October 2001

*Ingemar Eckerlund*

# Contents

# Overview and summary of the essays

# 1. INTRODUCTION

## 1.1 Background

Medical practice variation is a well-known phenomenon. Large variations in hospital admission rates, average length of stay for different diagnoses, frequency of surgical procedures, etc, have been documented in several studies in the last thirty years. This kind of variation exists at all levels of care and in all medical specialities. Moreover, the existence of variations also seems to be stable over time and across national boundaries. Regional (geographic) variations - *small area variations* - have attracted most interest, but studies of international variation are also frequent (Ham, 1988; Folland&Stano, 1989; Andersen&Mooney, 1990; Phelps, 1992; Mooney, 1994; Roos&Roos, 1994; Ashton et al, 1999; Wennberg, 1999).

The main part of the literature on medical practice variations derives from North America and Western Europe, including a large number of studies from the Nordic countries. Reports on this kind of studies have received considerable attention in health policy debates and, they have often been followed by hopes that the variations will indicate opportunities for savings or efficiency gains. This is not unreasonable, since many studies show that medical practice variations are uncorrelated with differences in morbidity or outcome of care (Wennberg, 1984; Andersen&Mooney, 1990; Ryan&Mooney, 1991; Phelps&Mooney, 1993).

However, it is far from self-evident that variation indicates inappropriate care. Particularly, high-use rates are not necessarily equivalent to an inappropriate overuse of services and low-use rates not necessarily equivalent to under-utilisation (Chassin, et al, 1987; Roos&Roos, 1994). Variations per se tell us nothing about the potential for enhanced efficiency. If the distribution as a whole is "off target", a small variation may create a much greater problem than a large variation where the distribution is generally "on target". Narrowing the variations might result in greater inefficiency, depending on where efficient practice lies in the range (Mooney, 1994). No particular level is a priori "right" under all circumstances, and some variation is of course desirable - provided that it reflects differences in medical needs and patients` preferences.

## 1.2 Purpose

The main purpose of this thesis is to show how economics can contribute to a better understanding of medical practice variations - why they exist, their consequences, and the extent to which they can – and should – be influenced. Greater insight into these issues is necessary before we can draw reliable conclusions on the potential savings and efficiency gains that may be concealed behind the variations.

A further purpose is to discuss how analyses of medical practice variations can be integrated into a framework of continuous performance monitoring, in order to promote efficiency in a systematic way. Finally, the ambition is to highlight the role of the empowered patient in the process of changing the practice patterns.

In principle, economics can contribute in two ways. First, by "positive economics", i.e., economic models and analyses that may help us understand the **causes** (including, but not limited to, economic factors) of existing medical practice variations and predict their **economic consequences**. Second, by "normative economics" (welfare economics), i.e., analyses of what policies are desirable and efficient to create **changes** in the actual state of the

world, e.g., to reduce the variation. This approach demands a discussion of "optimal variation", and the costs and benefits of reducing the variation.

The basic hypothesis is that medical practice variations to some extent indicate inefficiency in the utilisation of health care resources and consequently, a welfare loss to society. This hypothesis can be confirmed or rejected only when we have thorough knowledge about the underlying causes and consequences of variation. The causes may be influenced to a greater or lesser extent. Furthermore, the specific causes and consequences of the variation in question determine whether intervening against the causes of that variation is important or worthwhile, and the appropriate measures of intervention that should be used.

Given a certain case of variation (e.g., in cesarean section rates among hospitals), a natural first step is to formulate a theory about the causes and then build a model in order to test the theory/hypothesis. If the variation is explained by a variation in medical needs or patients' preferences, there might not be much to say. Otherwise, i.e., if the causes are less "legitimate", the next step is to find out what – if anything – can and should be done. This depends on the causes as well as the medical, economic, etc, consequences of the variation, and also on the opportunities for change. The more serious or costly the consequences are, the more important is it to take action, provided that something can be done at a reasonable cost. In other words, the measures considered must be evaluated with regard to their costs as well as their benefits.

# 2. MEDICAL PRACTICE VARIATIONS – PREVIOUS STUDIES

## 2.1 Causes

Previously, attempts to analyse the causes or determinants of medical practice variations were mainly made in the field of clinical epidemiology. Differences in morbidity, demography, availability or supply of resources, clinical judgement, patient expectations or demand, prevailing custom, as well as inaccuracies in data sources and random variation, are examples of proposed explanatory factors (Wennberg, 1984; McPherson, 1990; Wennberg, 1990). Systematic variation (i.e., excluding measurement errors and random variation) may be related to:

- *patient characteristics* (age, gender, morbidity, expectations, propensity to seek care, etc),
- *availability of resources* (supply of physicians, level of care, etc), and
- *policy* (tradition, practice style, etc).

Using economic terminology, the first two groups are demand-side and supply-side factors, respectively. The third group concerns supply as well as demand. On the one hand, established practice style is a specification of supply. On the other hand, it could be regarded as part of demand – the (secondary) demand generated by the physician/agent on behalf of the patient/principal. From the point of view of an economist, factors like price and incentives, which relate to the mechanism of resource allocation, should also be considered.

Which factors are most important in explaining existing variations depends on what organisational level is in focus, the level of aggregation of the variable studied and, naturally, the variability of the different explanatory variables. Despite numerous empirical investigations of small area variations, there are relatively few multiple regression studies. Many of the key variation studies are either univariate (using Chi-square tests for variations across market areas) or bivariate (tests of Pearson correlations of use rates with different variables). Although these methods are appealingly simple, they disregard that the problem is inherently multivariate, since the variation in use rates is, in principle, an issue involving simultaneous contributions of many factors.

A review of the literature on small area variations (Folland&Stano, 1990) shows that available multiple regression studies typically explain moderate to substantial amounts of the variation by standard socio-economic variables. However, often when a study explains a large portion of the variation by such variables, it has focussed on aggregates of procedures. Regression analyses at the small area level for individual procedures, on the other hand, have generally been less successful in explaining the variations.

Based on their literature review, *Folland&Stano* (1990) presented a number of recommendations for improving the small area variation studies. Thus, they suggest that future studies should incorporate multivariate approaches, and that the econometric equations should be specified as completely as possible. Further, the problem of endogeneity must be addressed, since a failure to do this may lead to causation being incorrectly ascribed to supply variables. They also point to the importance of using appropriate multicollinearity diagnostics, within the context of multivariate approaches, in order to avoid Type II errors (not rejecting the null hypothesis that there is no variation, when in fact it is false).

Even taking into consideration the methodological deficiencies of many previous studies, relatively equivocal experience indicates that patient-related factors generally contribute

relatively little to explaining existing variations for single procedures at the hospital (clinical department) level. In other words, the demand-side variables - besides age and gender, which are usually adjusted for - seem to have limited influence, probably because they do not differ much across hospitals (departments).

Regarding supply-side factors, two major arguments exist about their influence on medical practice variations - the ideas of "availability" and "inducement". The availability of hospital beds and physicians sometimes explains part of the variations, e.g. in hospital admission rates or rates of ambulatory surgery (Wennberg, 1984; Ham, 1988; Evans, 1990; Phelps, 1992). The types of insurance- and payment-systems may be important explanatory factors, mainly in countries with many different systems, and in international comparisons. In a study of the effect of Medicaid fee differentials on the use of cesarean delivery over the period 1988-1992, *Gruber et al* (1999) found that larger fee differentials between cesarean and normal childbirth for the Medicaid programme resulted in higher cesarean delivery rates.

According to available studies, supply-side factors are more important in the United States than in England or Sweden, which probably reflects differences in the health care systems. "Supplier-induced demand", i.e., demand provided in the self-interests of providers rather than solely for patient interests, has been a topic on the health economics research agenda, mainly in the United States (Evans, 1990; Ryan&Mooney, 1991; Labelle et al, 1994).

A related issue is that of the potential impact of malpractice fears on practice style, i.e., whether physicians´ malpractice fears lead to defensive medicine. Obstetrics is a speciality where defensive medicine is considered to be particularly problematic. Earlier studies have reported mixed evidence on the effects of liability fears on the probability of cesarean section. *Tussing&Wojtowycz* (1997) have shown that fear of malpractice influences the choice of a cesarean section both directly and through the use of electronic fetal monitoring, a major diagnostic tool, which is shown to influence the diagnosis of fetal distress. They conclude that "whether this is a desirable or undesirable effect remains ambiguous, but it is costly". In a later study, however, *Dubay et al* (1999) found that physicians practice defensive medicine in obstetrics but that the impact on total obstetric care costs of increased cesarean section from malpractice fears is small.

Often, however, the most important explanation seems to be variations in clinical judgement or practice style, the same kind of patients are treated differently in different hospitals (Phelps&Mooney, 1993). Leading researchers in clinical epidemiology have proposed that this is due to "professional uncertainty", i.e., there is a substantial uncertainty in diagnosing many conditions, and also an uncertainty associated with outcomes of alternative interventions (Folland et al, 1993; Wennberg, 1984). If there are several alternative methods for diagnosis and/or treatment of a certain disease, variations may be quite natural. On an objective, scientific basis, identifying a certain method as "best practice", a gold standard under all circumstances, may be impossible. This uncertainty naturally results in a lack of agreement about appropriate standards of care which, in turn, means that a physician's "beliefs" will partly determine his/her actions.

*Folland&Stano* (1989) have developed an econometric model of physicians´ resource utilization, including uncertainty and practice style as explanatory factors behind existing variations in the use of medical and surgical procedures. Practice style is modelled as an exogenous set of physician *beliefs* about the relationship between medical care and health status, i.e., about the health production function. Thus, practice style is distinguished from

supplier-induced demand, which is regarded as an endogenously chosen level of physician influence on patient demand. The fact that physicians' beliefs differ is assumed to be due to professional uncertainty. In the model, per capita utilization is decomposed into a first-occurrences demand and an intensity demand. It is assumed that practice style primarily affects the intensity with which the physicians treat their patients. Based on empirical tests of the model, the authors conclude that practice style does not seem to explain variations in resource utilization at an aggregated level, but would probably do so for single procedures.

The "professional uncertainty" hypothesis has been predominant. It has been confirmed by a number of studies, demonstrating that the variation, e.g. in hospital admission rates, tends to be smaller when there is a professional consensus on what constitutes best practice (Ham, 1988; Ryan&Mooney, 1991; Phelps, 1992). According to this hypothesis, medical practice variations may be more or less rational, but if the variations are due to lack of information, knowledge or professional competence, there may be room for enhanced efficiency.

The "professional uncertainty" theory has been questioned, however. According to the critics, the abundant evidence of existing variations would stimulate research and development, and the uncertainty should gradually vanish if this were correct. This does not seem to be the case, however. The pattern of variations is relatively stable over time, even for many well-established medical technologies. Thus, an alternative theory, the "professional certainty" hypothesis, has been launched. According to this theory, every doctor is perfectly certain that his/her practice style is correct, notwithstanding how much it differs from that of others (Evans, 1990; Ryan&Mooney, 1991; Chassin, 1993). This phenomenon is also known as "surgical signatures", reflecting the idea that the rates in markets as well as the surgical patterns across markets are stable and seem to remain constant over time (Wennberg&Gittleson, 1973). A closely related theory is the "enthusiasm hypothesis", suggesting that medical practice variations are caused by differences in the prevalence of physicians who are enthusiasts for particular services. This enthusiasm may also lead some to use these services for inappropriate reasons because they overestimate the utility of the procedure (Chassin, 1993).

Obviously, practice style also depends on the physicians' propensity to take calculated risks which, in turn, largely depends on clinical experience, competence and personal attitudes. Consequently, there are divergent - more or less well-founded - attitudes and opinions among physicians concerning what is best practice, i.e., *what* should be done, as well as *when* and *how*, *where* and *by whom*. It follows that the existence of different opinions on what constitutes "best practice" is not necessarily equivalent to uncertainty, and that there is a need for some kind of incentives - besides information - to influence and move practice style towards "best practice".

In a recent article, *Eldenburg&Waller* (2001) present a decision-case mix model for analysing the variation in cesarean section rates. The model was applied to a sample of deliveries at a hospital where physicians exhibited a considerable variation in their cesarean section rates. Comparing groups with a low versus high rate, the authors conclude that the difference in physician decision tendencies to perform a cesarean section, in response to specific obstetric factors, is at least as important as case mix in explaining variation in cesarean section rates.

*Phelps&Mooney* (1993) reviewed different "explanations" of variations from an economic perspective. Besides economic demand-side and supply-side factors, they discuss differences in illness patterns as well as random noise, physician differences and substitution in

production. Concerning the last explanation, they make a distinction between *substitution between inpatient treatments* and *substitution between inpatient and outpatient care.* Based on empirical results, they conclude that the extent of geographic variations can only be partly explained by standard economic phenomena such as price, income and the distributions of illness across regions. Substitution in production of health does not seem to occur very frequently. Rather, they systematically found a positive correlation among procedures that could medically serve as substitutes. Finally they turn their attention to "differences in beliefs about the efficacy of treatment and decisions about which patients should receive treatment", i.e., what has been called differences in practice style.

**2.2 Consequences**

Medical practice variations may have medical as well as economic consequences, and also equity consequences. Unfortunately, these consequences are not very well understood. Actually, there is a serious lack of knowledge about the medical consequences of different diagnostic and therapeutic methods and, consequently, a great need for medical technology assessment and outcomes research (Wennberg, 1990).

The same is true for the economic consequences of medical practice variations. The literature on this issue is very limited. One reason for this scarcity may be what has been characterised as a tendency among economists "to dismiss the variations literature as irrelevant, or else briefly to explain it away" (Phelps&Mooney, 1993). Most evaluations of the economic consequences of practice variations have been more or less crude estimations of "potential savings given that all units produce the minimum rate of a certain surgical procedure", and the like. Not only is that kind of calculations dubious from a methodological point of view, since nothing says that the minimum rate is the correct one, in most cases they also give an incorrect impression of the true savings potential.

However, in an article on "Priority Setting for Medical Technology and Medical Practice Assessment", *Phelps&Parente* (1990) use the tools of demand theory to analyse the economic consequences of medical practice variations in terms of welfare loss. A basic assumption in their model is that, as the indications for using a certain medical technology are expanded, it will display decreasing marginal utility. They show that the total welfare loss to society, resulting from interarea deviations from the correct rate of use of a certain medical technology, is a function of the total expenses for that technology, the coefficient of variation of inappropriate use, and the absolute value of the price elasticity of demand.

In the early 1990's, Journal of Health Economics was the forum for a theoretical debate on the valuation of the costs of the waste associated with health insurance. *Rice* (1992) challenged the traditional method - developed by Feldstein and used by *Feldman&Dowd* (1991) - to calculate the welfare loss of "excess insurance", asserting that consumer ignorance (asymmetric information) and the prevalence of substantial amounts of unnecessary care make the demand for care a poor basis for such calculations. The question in focus was whether demand reflects the true valuations (utility) of the consumers (patients). This debate was also of great relevance for a discussion of the economic consequences of medical practice variations, and included a debate between *Dranove* (1995) and *Phelps* (1995) on the validity of the above-mentioned Phelps&Parente model.

Recently, *Nyman* (1999) initiated a new debate on the value of health insurance in the same journal. He argues that in the existing literature, the gains from insurance have been underestimated while the efficiency losses from moral hazard have been overestimated.

Further, he argues that these losses "may be outweighed by the welfare gains derived from the income transfers that allow the consumer to consume more medical care and other commodities in an ill state, or that simply allow the consumer to gain access to medical care procedures that would otherwise have been unaffordable." *Blomqvist* (2000) and *Manning&Marquis* (2000) answered Nyman's article and the debate is probably not yet over.

Without ending up in consensus, these debates clearly indicate that the issue of economic consequences of medical practice variations very much centres on the economics of information, e.g. problems of uncertainty and asymmetric information.

**2.3 Opportunities for Change**

Among the causes of medical practice variations discussed above, the demand-side factors could, more or less, be regarded as legitimate. They partly describe facts, like age, gender and morbidity, partly patient preferences. Consequently, the factors that can (should) be influenced are the supply-side factors. Some of these, like resource capacity (staff density, bed supply, etc) may, at least in principle, be relatively easy to influence, e.g. by budget restrictions. However, the supply-side factors reflecting policy and practice style are probably more resistant to change. Therefore, the focus should be on these factors and particularly on appropriate incentives for change.

Experience shows that it is difficult to influence practice by information alone. Continuous feedback of information on actual practice variations is a necessary, but not a sufficient, prerequisite for influencing practice (Evans, 1990). However, reports on actual variations in combination with discussions about economic concepts like *opportunity cost*, and tentative calculations of potential efficiency gains, have been shown to stimulate self-assessment and change of medical practice (Eckerlund et al, 1992). Some studies have shown that educational, feedback, and surveillance programs directed at physicians can alter their behaviour and thus, presumably, their practice style (Dyck, 1977; Wennberg et al, 1977; Nyman et al, 1990). Experience from using practice guidelines as a means of changing practice patterns is mixed (Goodpastor&Montoya, 1996). For example, a widely disseminated and nationally endorsed practice guideline for cesarean surgery had almost no impact on the pattern of obstetrical practice in Ontario, Canada (Lomas, et al, 1989).

Improved knowledge and information about the economic consequences (costs) of different medical practices would probably be a valuable basis for influencing practice. Another prerequisite is knowledge about what constitutes "best practice", taking into account medical as well as economic and other effects. As a matter of fact, many methods and routines frequently used in the health care sector lack scientific support (Wennberg, 1990) and thus, medical technology assessment should be a high priority area.

Based on a literature review concerning various attempts to influence medical practice, *Greco&Eisenberg* (1993) describe six general methods: *continuous education, feed-back of information, participation by physicians, administrative interventions, financial incentives,* and *financial penalties*. The authors conclude that a combination of methods is more likely to result in permanent changes than any single method. Before deciding on a certain action plan to change practice, one should try to find out whether the intended program is appropriate, supported by the physicians, and whether it will be perceived as a threat or an opportunity for improvement. An active involvement and support from those physicians who will be affected is an important prerequisite for success.

From a change perspective, it is interesting to note that, in recent years, greater emphasis has been put on the patient's individual autonomy in medical decision making, i.e., to let patient preferences influence the choice of diagnostics and treatment strategies. In most countries, there is an increasing emphasis on shared information as well as shared evaluation, shared decision making, and shared responsibilities (Coulter, 1999). She states that "patients have grown up – and there is no going back".

The same is probably also true for letting patients and citizens have a say about the quality of health services, in order to use their views and experiences for improving the quality of care. Various methods have been applied for measuring patient satisfaction (Nathorst-Böös et al, 2001), but there is great need for further development of methods, particularly methods including health economic aspects.

The focus on the "empowered" patient has also led to a re-examination of the nature of the agency relationship in health care. For example, *Vick et al* (1998) present a study, based on an econometric (random effects probit) model, concerning the relative value to patients of various attributes of the general practice consultation. They found that the most important attribute was "being able to talk to the doctor", whilst "who chooses your treatment" was the least important. Only females and highly qualified respondents preferred to choose their treatment themselves.

The physician-patient encounter is also discussed by *Gafni et al* (1998). Assuming a goal of arriving at a treatment decision based on the physician's knowledge and the patient's preferences, they discuss the feasibility of two alternative treatment decision-models: (1) the physician as a perfect agent for the patient, and (2) the informed patient treatment decision-making models. They argue that transferring information to the patient is easier and less resource demanding than transferring each patient's preferences to the physician in each medical encounter. Accordingly, they find that designing contracts to motivate physicians to transfer information to patients is more feasible than designing contracts to motivate physicians to find out their patients' utility functions. In a comment to Gafni et al, *Rochaix* (1998) argues that, beyond the discussion of the relative merits of the two models, two features need to be integrated for an adequate account of this complex interaction: the two stages of the physician's action (the diagnosis and the treatment) and the physician's double agency (to the patient and to the representative of the collective interest).

# 3. THEORETICAL FRAMEWORK

## 3.1 Medical practice variations and economic models

Traditionally, economists have modelled self-employed physicians as owner-managers of business firms. A model presented by *Zweifel&Breyer* (1997) provides a good illustration of this approach. The authors assume that the physician's utility depends on three factors only: the physician's annual net income (Y), the hours of work that the physician must devote to medical practice ( t), and the degree of "artificial demand creation" practised by the physician (s). The physician's utility function is expressed by the equation:

$U = U(Y,t,s)$, with first derivatives $U_Y>0$, $U_t<0$, $U_s<0$,

where $U_Y>0$ indicates the hypothesis that, everything else equal, the physician's utility increases with income. $U_t<0$ expresses that, everything else equal, physicians prefer spending more time on activities other than practising medicine and that they devote time to medical practice only because it earns them an income. The third derivative, $U_s$, relates to physician-induced demand. The stated hypothesis, $U_s<0$, implies that, everything else equal, artificial demand creation always gives the physician bad conscience. An $U_s=0$ would mean that the physician is not troubled at all by artificial demand creation.

It is assumed that physicians will manage their time and conduct their practice so as to maximise their own utility, subject to the following constraints:

- the market for their services
- the markets for office personnel, supplies, equipment, floor space, and so on
- the technological state of the art of medicine
- the government regulations
- their own conscience regarding demand creation, and
- the fact that only 24 hours a day can be split between time in their practice and time spent on other pursuits.

This theoretical model implies that, at any given level of demand creation, physicians will always behave so as to maximise the hourly net income they can extract from the practice of medicine. Furthermore, the model implies that the physician will naturally create artificial demand for his or her services, to the point at which the marginal benefit of the additional income earned equals the marginal utility loss of the additional working hours and the bad conscience resulting from demand creation.

Thus, ethical conduct in this model emerges from a trade-off among income, the desire for more leisure time, and the physician's conscience, and it varies with the economic circumstances. Physicians are not seen adhering rigidly to a professional code of ethics (Reinhardt, 1999).

The relevance of this theoretical model can, naturally, be questioned, especially in a health care system like the Swedish one, where most physicians are salaried by the county council and not working on a fee-for-service basis. The power of this model at the empirical level has been limited, even in health care systems with a larger number of self-employed physicians.

In recent years, a number of articles based on economic models of variation in cesarean section rates have been published in health economic journals, of which one of the most interesting is the one by *Chetty* (1998).

Chetty states that the decision problem for physicians involves choosing a treatment or procedure for a medical condition, based on the observed health status of patients. A correct treatment can improve health, while an inappropriate treatment can do some harm., i.e., input in this production process produces "goods" as well as "bads" with some statistical regularity. An important feature of any statistical decision problem including the physician's is that a reduction of one type of error is only achieved by increasing another type. Input costs are not only the direct costs of treatment, but also the costs of unnecessary treatments. Thus, the economic problem as stated by Chetty is: what combination of errors to produce, when production possibilities of correct decisions are constrained, and how to produce and for whom.

Chetty uses a model including patients, physicians, insurers and uncertain diagnostic technology to derive the optimal cesarean rate from preferences, technology and the incidence rate, when the choice of insured patients is only constrained by technology.

It is demonstrated that uncertain diagnoses produce unnecessary cesareans and unsafe vaginal births. Technical progress might lead to more cesareans and higher costs. According to this model, joint production of goods and bads and collective payments require incentive compatible pricing schemes, different from the Resource Based Relative Value Scale (RBRVS). Equilibrium outcomes of HMO's and fee-for service organisations are shown to be identical. However, implementable incentive schemes involve additional costs. Efficiency requires insurers, and not providers, to be liable for malpractice claims.

### 3.2 Medical decision-making, agency and practice style

To further understand medical practice variations, it is important to understand medical decision-making, i.e., what governs the choice of diagnostic, therapeutic and other procedures. Various theories have been launched in the literature concerning the major determinants of medical decision-making, that is, the physicians` practice behaviour. Not surprisingly, most theories are derived from a medical perspective.

Based on an extensive review of the literature, *Eisenberg* (1986) describes three groups of factors, related to *the self-fulfilling physician*, the physician's role as *the patient's agent*, and the physicians's desire to *maximise the social benefit* of medical care. The first group relates to the physician's economic self-interest, personal style, and practice environment. The second group relates to the physician's desire to act on behalf of the patient's physical or economic health and the patient's preferences. The third group, finally, is based on the idea that the physician is guided by a desire to provide the most good to the most people.

In economics, the doctor-patient relationship has frequently been modelled in a theoretical context based on contract theory or, more exactly, the economic theory of agency (Hart&Holmström, 1987; Blomqvist, 1991; Scott, 1996). The principal-agent theory is characterised by an agent having to choose an action on behalf of the principal. The action influences the welfare of both the principal and the agent, who are assumed to maximise their respective utility functions subject to the constraints they face. The relation is also characterised by asymmetric information, i.e., an imbalance of information between the principal and the agent, typically in favour of the agent. This asymmetry of information takes

two forms; *hidden action* (moral hazard) and *hidden information* (adverse selection). Hidden action means that the principal cannot observe the actions (efforts) of the agent, but only the outcome of those actions. Hidden information means that the agent has information about the action that the principal has not, and that although the action (effort) may be observable, the principal does not know whether it is the most appropriate.

The principal must devise a contract that is *incentive compatible*, i.e., that provides incentives for the agent to maximise the principal's (expected) utility, and it must ensure some minimal specified level of (expected) utility for the agent. Furthermore, this minimal level of utility for the agent must, at least, equal that achieved by alternative contracts with other principals. This is the *participation constraint.*

The patient-physician relation seems to be a good example of the principal-agent relation. The patient (principal) is unable to monitor the physician's (agent's) efforts and the relation between effort and outcome is uncertain. However, there are several differences between standard agency theory, and agency theory applied to health care (Mooney&Ryan, 1993; Scott, 1996).

While the standard theory assumes one principal and one agent in direct contact, there are many different types of agency relationships in health care. Besides the patient-physician relationship, we could think of agency relationships between the patient and the third party payer, between the physician and his/her trade union, and also between the third party payer and the physician (or his/her trade union). Although the patient-physician relationship is the one in focus, all these relationships are interdependent (Evans, 1984). The patient-physician relationship is influenced by the others, and instead of being a perfect agent for the patient, the physician can be expected to act as a "double-agent" - for the patient as well as for society (Blomqvist, 1991).

While the utility functions of principal and agent are independent in the standard theory, the physician's and patient's utility functions are, to a certain extent, interdependent in the health care setting. We assume that the physician's utility function includes the patient's utility as an argument (Evans, 1984). This implies that physicians want to "do their best" for the patients and, consequently, that "effort" is positively related to the physician's utility, maybe with diminishing marginal utility. Rewards/penalties in health care may take on social as well as monetary form, while only adopting financial form according to the standard theory.

In the standard theory, the principal's utility depends only on income, part of which is the fee paid to the agent. Assuming that both parties are risk averse, some degree of risk sharing is optimal. In health care, on the other hand, the arguments in the patients' utility function are non-pecuniary, e.g. health status, which alters the nature of risk-sharing. The patient bears the risk of the health outcome being less or more than expected, the physician bears the risk of a poor outcome only in case of malpractice. If this happens, there may be some risk in terms of loss of reputation, but not necessarily income. However, it is still the case that risk averse physicians will change their behaviour in response to the perceived probability that they may be sued, i.e., they may engage in defensive medicine (Mooney&Ryan, 1993; Scott, 1996).

Finally, concerning the "participation constraint", standard theory assumes that agents have alternative contacts from which to choose. In Swedish health care, however, the alternatives are very limited. Furthermore, the Swedish health care system, with hospital-employed physicians and specially appointed chief physicians responsible for all health care at their

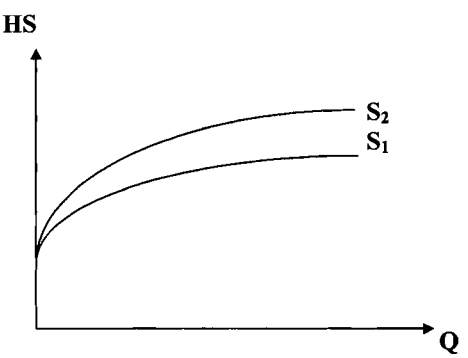
departments, makes it reasonable to assume a consistent practice style within the single department. This, in turn, means that the relevant study unit is the clinical department rather than the single physician (Westert, et al, 1993).

To summarise, the medical decision can be described as an optimisation problem under uncertainty. The physicians/departments are supposed to maximise a utility function including their own utility as well as the patients' utility and social utility. The utility maximisation is constrained by a variety of restrictions - medical, organisational, technological, economic, etc.

Theoretically, the solution to this optimisation problem defines the demand for health services which, in combination with "supply", results in the actual use of health services. Our basic hypothesis is that the existing medical practice variations are, at least partly, due to inefficiency - production as well as consumption inefficiency. In our empirical studies, we define a number of variables, which are assumed to influence the demand and supply of health care. On this basis, we specify regression models to test our hypothesis. Naturally, it is very important to base these models on reasonable assumptions about the arguments in the patients' and physicians' utility functions.

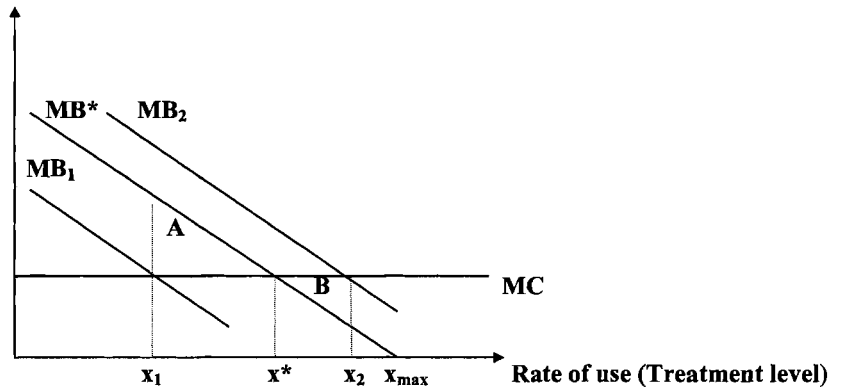
Much of our information suggests that medical practice variations are largely due to differences in physicians' practice style. In figure 1, the intensity of care (Q) is measured along the horizontal axis and the patient's health status (HS) along the vertical axis. Curves $S_1$ and $S_2$ represent the different beliefs of two physicians, where physician 2 believes that additional units of service have larger beneficial effects (Folland&Stano, 1989).

**Figure 1** Physician's beliefs about the relationship between medical care and health status




Different beliefs about the production function give rise to different marginal benefit curves (see Figure 2).

**Figure 2 Marginal benefit and marginal cost as functions of use rate**




Assuming that MB* is derived from the "true" production function, the optimal treatment level is x*, where marginal benefit, MB, equals marginal cost, MC. The treatment levels given by the intersection between $MB_1$ ($MB_2$) and the MC-curve , $x_1$ ($x_2$) thus indicate under-utilisation (over-utilisation). The welfare losses are represented by triangles A and B, respectively. If the patient has complete information and pays for the care out-of-pocket, then she will demand x*. If, instead, the patient is fully insured (which, in principle, is the case in Sweden), s/he will maximise well-being by demanding $x_{max}$. A physician acting as a perfect agent for the patient would also provide or recommend this quantity ($x_{max}$) of care. However, in the Swedish setting, the physician should probably be regarded as a "double-agent" for several principals (the patient, the hospital/department, the County Council and the community), with at least partially conflicting interests. A probable compromise solution would be somewhere between x* and $x_{max}$, which means that there is an incentive towards over-utilisation.

Besides, there are other incentives working in the same direction. From the physician's perspective, two kinds of mistakes (incorrect decisions) are possible: to perform a certain procedure when it is "inefficient" (to the right of x* in figure 2), and not to perform it when it is "efficient" (to the left of or exactly at x* in figure 2). Both kinds of mistakes may have medical as well as economic consequences. However, the consequences are not the same, and attempts to reduce the first kind of mistake may increase the probability of the second, and vice versa. The incentives to avoid the second kind of mistake, and thereby reducing the probability of malpractice litigation, seem to be much stronger than the incentives to avoid the first kind of mistake, which means that there is a built-in bias towards over-utilisation.

### 3.4 The cost of inefficient variations

From an economic point of view, the important issue concerning medical practice variations is whether they indicate efficiency problems, i.e., whether it is possible - through reallocation and alternative utilisation - to gain more utility from available resources. In economic literature, it is common to distinguish between two kinds of efficiency problems - technical (production) inefficiency and allocative (consumption) inefficiency, respectively.

The first kind of inefficiency simply means that a commodity is produced in an incorrect (wasteful) manner. Thus, it would be possible to produce more with available resources or,

alternatively, achieve the actual production with less resources. Variations in the average length of hospital stay for the same kind of patients (diagnosis group) might indicate production inefficiency, for example some hospitals might use more resources (bed-days) than necessary to produce a certain output.

The second kind of inefficiency means that the mix of commodities produced is incorrect (inappropriate). In health care, consumption inefficiency may occur as over-utilisation or under-utilisation. In the former case, the actual method (procedure, treatment) is utilised beyond the point where the marginal benefit to the patient (society) is out-weighted by the marginal cost, whereas in the latter case, its potential benefits are not fully exploited. Obviously, utility gains are possible through reallocation of resources from the former to the latter type of applications.

Naturally, both kinds of inefficiency may coincide, e.g. when too much of a certain procedure is carried out, and in a wasteful manner. Total economic efficiency presupposes production efficiency as well as consumption efficiency which, in turn, presupposes that producers and consumers face the same relative "prices". The following discussion mainly focuses on the problem of consumption inefficiency.

*Phelps&Parente* (1990) showed that the total welfare loss (W) to society resulting from interarea deviations from the correct[1] rate of use of a certain medical procedure, i.e., the potential gain from eliminating undesirable variation, is equal to the sum of all triangles A, B, etc (Figure 2). That sum is given by the following formula:

$$\mathbf{W = 0.5 \times (Total\ expenses\ for\ X) \times (COV)^2 / \eta},$$

where COV is the coefficient of variation of inappropriate use (i.e., the standard deviation in "unexplained" use divided by the average use of X)[2] , and $\eta$ is the absolute value of the price elasticity of demand. Given that the demand curve is known, and thereby $\eta$, W can be interpreted as the money value of the information needed to get all departments to adjust to MB*(x), i.e., to the best practice.

The formula implies that the welfare loss increases when large numbers of patients are affected, the average cost of intervention is large, the level of uncertainty about the correct use rate is high, and the marginal value falls rapidly as the rate of use expands. Assuming constant marginal costs, we can use the average cost of the procedure to calculate the relevant total spending. The remaining problem is to estimate the marginal benefit (demand) curve, and thereby $\eta$. *Phelps&Mooney* (1992) recommended $\eta$ =0.15 as a global estimate of demand elasticities, obtained from the RAND Health Insurance Study, HIS (Manning et al, 1987). Unfortunately, we have no corresponding estimate of $\eta$ in Sweden. There is no easy way of solving this problem, but we might be able to use a contingent valuation approach, i.e., ask patients about their willingness-to-pay for (the valuation of) certain medical procedures under specified conditions. As a provisional arrangement, we could carry out a sensitivity analysis based on fictitious values of the elasticity of demand.

---

[1] Phelps&Parente first assumed the average amount of any medical intervention (X) across all regions to be the correct one, and then showed that this assumption probably implies an underestimation of the true welfare loss.

[2] This "modified $COV^{2}$" = $(1-R^2)\ COV^2$.

# 4. SUMMARY OF ESSAYS I-V

The common purpose of the essays is to contribute to a better understanding of medical practice variations – why they exist, their potential consequences, and to what extent they can and should be influenced. The table below indicates the focus of the five essays on causes, consequences and opportunities for change, respectively.

| | Causes | Consequences | Opportunities for change |
|---|---|---|---|
| Essay I | X | X | |
| Essay II | | X | |
| Essay III | X | X | X |
| Essay IV | | | X |
| Essay V | | X | X |

The first essay, *Econometric analysis of variation in cesarean section rates: a cross-sectional study of 59 obstetrical departments in Sweden*, is an attempt to identify the causes of variation in cesarean section rates and discuss their economic consequences. The second essay, *Estimating the effect of cesarean section rate on health outcome – evidence from Swedish hospital data*, is based on cross-sectional data from the same 59 departments for five years, and its objective is to find out whether the cesarean section rate has a positive effect on health outcome. The third essay, *Benchmarking in obstetric care: a comparative study based on Data Envelopment Analysis*, focuses on variation in productivity in inpatient obstetric care among 19 Swedish hospitals, but also on strengths and weaknesses of the units studied, and the potential for enhanced productivity.

The last two essays deal with a new method for incorporating patient perspectives into the management of health care, i.e., as a basis for quality improvement. *Change-oriented patient questionnaires – testing a new method at three departments of ophthalmology*, is a pilot study using a method that is explicitly change-oriented. The last essay, *Patient satisfaction and priority setting – an economic approach*, is based on the same method but takes this approach one step further by analysing if, and how, priorities are influenced when an economic dimension is included.

### 4.1 Essay I

The purpose of this study is to explain the variation in cesarean section rates among obstetrical departments in Sweden and discuss the potential economic consequences of this variation. Based on a single year cross-sectional sample of 59 obstetrical departments in 1991, we identified some 20 determinants, demand-related as well as supply-related. A general, unrestricted, linear regression model including all these regressors was specified. After reducing this model, we were able to explain about 27 percent of the interdepartmental variation.

According to our final model, the cesarean section rate increases with the percentage of mothers over 35 years, and the percentage of cases with indication *placenta praevia*, but decreases with the existence of a systematic follow-up of performed cesareans. Even within the different indication groups we found a large variation in cesarean section rates among departments. This indicates that the decision on mode of delivery is, to a large extent,

governed by *practice style* which, in turn, is assumed to reflect the attitudes and beliefs of the obstetricians/departments concerning the health production function. The variation in cesarean section rates within the various indication groups is probably not only due to differences in the probability of cesarean section, given a certain indication (*treatment practice style*), but also to differences in the propensity to set a certain diagnosis (*diagnosis practice style*).

We conclude that the unexplained variation in cesarean section rates indicates inefficiency, mainly due to over-utilisation. Our discussion concerning the economic consequences results in some rough estimates, showing an *additional cost* for "unnecessary" cesarean sections of 13-16 million SEK per year. The *welfare loss* to society due to undesired variation is tentatively estimated to be about twice as high.

### 4.2 Essay II

While, in the first study, we implicitly assumed that there was no difference among departments in health outcome, this paper tests the null hypothesis of a zero effect of the cesarean section rate on health outcome, against the alternative of a positive effect. Besides the traditional outcome measure, *perinatal mortality*, we also used a morbidity measure, *rate of asphyxia*.

An obstetrical department may have a high cesarean section rate because of a high prevalence of risk cases or complicated deliveries, which makes it necessary or desirable to perform a relatively large number of cesarean sections. When controlling for case-mix, however, we should expect to find a *negative* correlation between outcome (perinatal mortality, rate of asphyxia) and cesarean section rate.

We used hospital data on health outcome, cesarean section rates and some further variables measuring "case-mix", from 59 obstetrical departments in Sweden 1988-1992. We specified two separate linear regression models, one with perinatal mortality, and the other with rate of asphyxia, as the dependent variable. Besides the cesarean section rate, we included the percentage of mothers over 35 years, the percentage of multiple births, and the percentage of (ten) different obstetrical indications as independent variables to control for case-mix. These models were estimated by single-year cross-section and pooled cross-section regressions.

We could not reject the null hypothesis of a zero effect, i.e., we did not find any significant positive relationship between the cesarean section rate and health outcome in terms of perinatal mortality or rate of asphyxia. Thus, we conclude that an increase in the cesarean section rate does not imply lower perinatal mortality or a lower rate of asphyxia. This, in turn, indicates that the minimal cesarean section rate is optimal. We end by stating that a natural next step would be to elaborate on the methods for estimating the welfare loss due to undesired variation, and also to identify and evaluate the policy options, i.e., the opportunities for change.

### 4.3 Essay III

Benchmarking is an important tool for self-assessment as well as external accountability to promote efficiency in the health care sector. However, a prerequisite is that the benchmarking process is based on scientifically sound methods. Using DRG-data from 19 Swedish hospital departments in 1994-1995, we apply Data Envelopment Analysis (DEA) to compare technical efficiency (productivity) in inpatient obstetric care. Output is measured in terms of cesarean

sections, vaginal deliveries and other DRGs, while physician costs, costs for other staff and number of beds are used as inputs.

The results indicate an average potential for enhanced efficiency of at least 6-7 percent each year. Most hospitals show a decreasing productivity in 1994-1995, partly due to the inability to adjust production capacity and costs to decreasing demand. The efficient units were found among departments of all sizes, while the inefficient ones were found among small and large departments.

From a policy perspective, an important advantage of the DEA method is that it helps identify strengths and weaknesses in the units studied. We conclude that there are good reasons for employing the DEA method for benchmarking in other speciality fields as well, and preferably on a larger scale, i.e., with more observations, and for including quality indicators among the output variables.

**4.4 Essay IV**

In recent years, various methods for measuring patient satisfaction have been applied as part of quality improvement programmes. The purpose is to incorporate patient perspectives into the management of health care. However, these survey methods have often been defective in terms of validity and reliability and with respect to the potential for using the results to improve the quality of health care. This essay presents a study at three departments of ophthalmology in Sweden, involving a new method, Quality Satisfaction Performance (QSP), that fulfils reasonable demands for validity and reliability, and is explicitly change-oriented. The parameters of the model are estimated with simultaneous equation methods, based on partial least squares technique.

Data were collected by a questionnaire survey at three ophthalmology departments in May and June 1994. Patients in three segments – cataracts, strabismus, and diabetes – answered questionnaires containing 49 questions about different quality dimensions at the departments. Approximately 1500 patients were invited to participate in the study and the final response rate averaged 80 percent.

The results show consistently high patient satisfaction indices. All departments got an index value between 86 and 90 (out of 100). Satisfaction varies somewhat among the different patient segments and departments. Cataract patients were found to be the most satisfied patient group at all three departments, and the lowest index for cataract patients (90.5) was higher than the highest index figure for the other groups. In addition to patients' rating of different quality dimensions, the QSP method also shows the impact of these dimensions on overall patient satisfaction. All patient groups in the three departments gave low ratings to accessibility, a quality dimension that was demonstrated to have a high impact on overall satisfaction, and thus a strong candidate for improvement. We conclude that decisions on improvements must be preceded by a thorough assessment of costs as well as the effects associated with the various changes.

**4.5 Essay V**

Several studies have shown that the QSP model can produce useful and policy relevant results. However, from a health economics perspective, the lack of economic content in the basic model is a serious deficiency. The purpose of this study is to analyse if and how priorities according to the QSP approach are influenced when an economic perspective is explicitly included. This is accomplished by a cost-effectiveness analysis of certain proposed

changes/improvements, a cost-benefit analysis based upon the patients' willingness-to-pay for these changes, and a correlation analysis of priorities based on patient satisfaction, willingness-to-pay, cost-effectiveness analysis and cost-benefit analysis, respectively.

Data were collected by a patient survey at the Department of Obstetrics and Gynaecology at Karolinska Hospital in Stockholm, Sweden. All patients visiting the clinic for the first time during the period August 1998 - February 1999 received a survey questionnaire including 43 closed response questions about their judgements of the quality dimensions accessibility, environment and participation, and their willingness to pay for proposed improvements. In total, 800 original questionnaires were distributed and the final response rate was 82 percent.

Our results show that the ranking between various improvements is strongly influenced when an economic dimension is included. We conclude that even a methodologically appropriate measurement of patient satisfaction may lead to a cost-ineffective priority setting, unless economic consequences are explicitly considered. The cost-effectiveness analysis gives a basis for priority setting and the cost-benefit analysis tells us how much should be spent on improvements.

# 5. DISCUSSION AND CONCLUDING REMARKS

The title of this thesis "...the economics of medical practice variation" is to be understood in a broad sense. The purpose is to show how economics can contribute to a better understanding of the variation phenomenon, which includes the role of economic analyses in performance monitoring and priority setting based on patient views.

Using cesarean section as an example, we specified and estimated an econometric model in order to explain the variation in cesarean section rates among hospitals. Since only a minor part of the variation could be explained by differences in the demand- and supply-related variables included, and since a cesarean section is more resource demanding than a vaginal delivery, we concluded that the variation indicates inefficiency. Next, in order to find the optimal rate, we tried to estimate the effect of the cesarean section rate on health outcome. We tested the null hypothesis of a zero effect against the alternative of a positive effect. The null hypothesis could not be rejected, i.e., we did not find any significant positive relationship between the cesarean section rate and health outcome.

Much of our information suggests that the unexplained variation is largely due to differences in physicians' practice styles. However, these are difficult to identify and even more difficult to quantify. Using data from the Medical Birth Registry makes it difficult to capture exactly what governs the decision to carry out a cesarean section or not. To find out more in detail which factors influence the medical decision-making, we probably have to rely on other kinds of data and analytical methods. Careful survey studies of a number of individual cases might enable us to learn more about the influence of practice style on medical practice variations. This seems to be an area in need of further research. Possibly, multilevel analysis techniques could help us separate out the effect of the practice style factor (Gatsonis et al, 1993; Westert et al, 1993).

Regarding the economic consequences of medical practice variation, we present some tentative and rough estimates. This is a field in great need of further research and development. The method developed by Phelps&Parente (1990) seems to be an interesting approach, suitable for our purposes. However, before we can fully adopt that method, we need to estimate the marginal value curve. Since there is no real *market* for health care in Sweden, there is no self-evident method for estimating demand. Possibly, we could use the contingent valuation method (Johannesson, 1996; Johansson, 1995) to that end, i.e., ask patients about their willingness-to-pay for a certain procedure, given specified risks and benefits, and thereby derive the "demand" – for cesarean sections as well as other procedures.

Concerning the opportunities for change, the literature review tells us that it is difficult to influence practice by information alone. Assuming that medical practice variations are largely due to differences in practice style, whether reflecting differences in physicians' attitudes and beliefs about the effectiveness of medical technology (professional uncertainty) or differences in enthusiasm, we have to find means and measures to motivate efficiency-enhancing changes. Probably, what is needed is a combination of methods, like continuous feed-back of information on performance and practice variation, clinical practice guidelines, education and appropriate incentives. To identify and implement the incentives necessary to move practice style towards best practice, we need more information and knowledge about the utility functions of patients and physicians. Furthermore, active involvement and support from those physicians who will be affected, is a prerequisite for success.

The benchmarking approach in obstetric care, based on data envelopment analysis (DEA), seems to be very useful from a policy perspective. Opportunities for enhanced efficiency were demonstrated and the analysis also indicated strengths and weaknesses of the different departments, which should be of great value when it comes to decisions on changes. There are good reasons for adopting the DEA method on a larger scale as a component of a system for continuous performance monitoring.

Patients' perceptions of quality are essential for determining effectiveness and efficiency in health care delivery, and should thus be utilised as a basis for decisions on quality improvements. We used the QSP approach, which has been demonstrated to produce useful and policy relevant results. However, from a health economics perspective, the lack of economic content in the basic model is a serious deficiency. To remedy that deficiency, we estimated the costs, as well as the effects, of various improvements. In that way, we are able to rank potential improvements according to their cost-effectiveness ratios. In addition to this, we also asked the patients to evaluate the benefits in terms of their willingness to pay for the changes, in order to rank the changes according to their benefit-cost ratios.

We found that the ranking between various improvements is obviously influenced when an economic dimension is included. We also found disagreement between the cost-effectiveness and the cost-benefit ranking in a few instances. This indicates a need for further analysis and model development. It also points at the importance of studying both the benefit and the cost side, when considering alternatives for possible changes.

Finally, it should be kept in mind that the mere fact that variations exist does not tell us what is right or wrong. The answer to the question "What is appropriate care?" would be that it depends. Partly, it is a task for medical technology assessment. But it also depends on the preferences of the patients and citizens, and the amount of available resources. Medical practice variations, if caused by genuine professional uncertainty, may indicate a need for technology assessment. Thus, there is a two-way link between studies of medical practice variation and technology assessment. In that respect, the economic consequences of medical practice variations may be an important basis for priority setting in technology assessment.

## REFERENCES

Andersen, T.F., Mooney, G. (eds). The Challenges of Medical Practice Variations. London: Macmillan, 1990.

Ashton, C.M., Petersen, N.J., Souchek, J., et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. New Engl J Medicine 1999;340;32-9.

Blomqvist, Å. The doctor as double agent: information asymmetry, health insurance, and medical care. Journal of Health Economics 10 (1991):411-22.

Blomqvist, Å. Does the economics of moral hazard need to be revised? A comment on the paper by John Nyman. Comment. Journal of Health Economics 2001;20:283-288.

Blomqvist, Å. Rejoinder. Journal of Health Economics 2001;20:299.

Blumenthal, D. The variation phenomenon in 1994. N Engl J Med 1994;331:1017-18.

Chassin, M.R., Kosecoff, J., Park, R.E., et al. Does inappropriate use explain geographic variations in the use of health care services? JAMA 1987;258:2533.

Chassin, M.R. Explaining Geographic Variations: the enthusiasm hypothesis. Med Care 1993;31:YS37-YS44, Supplement.

Chetty, V.K. Stochastic technology, production organization and costs. Journal of Health Economics 1998;17:187-210.

Coulter, A. Paternalism or partnership? Patients have grown up – and there's no going back. BMJ 1999;319:719-20.

Diehr, P. et al. What is too much variation? The null hypothesis in small-area analysis. Health Services Research 1990;24:741-771.

Dranove, D., White, W.D. Agency and the Organization of Health Care Delivery. Inquiry 1987;24:405-415.

Dranove, D. A problem with consumer surplus measures of the cost of practice variations. Journal of Health Economics 1995;14:243-251.

Dubay, L., Kaestner, R., Waidman, T. The impact of malpractice fears on cesarean section rates. Journal of Health Economics 1999;18:491-522.

Dyck, F.J. et al - Effect of Surveillance on the Number of Hysterectomies in the Province of Saskatchewan. N Engl J Med 1977;296:1326-1328.

Eckerlund, I., Håkansson, S. Variations in resource utilization - the role of medical practice and its economic impact. Soc. Sci. Med. 1989;28:165-173.

Eckerlund, I., Brege, K.G., Thorburn, W. Ögonsjukvården 1988-1991 : resultat och erfarenheter av utvecklingsarbete med kontinuerlig verksamhetsuppföljning. Stockholm : Spri, 1992. (Spri rapport 334). (In Swedish).

Eckerlund, I., Bengtsson, T., Gårdmark, S., Ljungblad, U. Diskussion och självrannsakan har lett till effektivare vård. Läkartidningen 1992;89:4169-70. (In Swedish).

Eckerlund, I. Economic impact of medical practice variations. Centre for Health Economics, Stockholm School of Economics, 1997.

Eisenberg, J. Doctor's decisions and the cost of medical care: the reasons for doctors' practice patterns and ways to change them. Ann Arbor, Michigan: Health Administration Press Perspectives, 1986.

Eldenburg, L., Waller, W.S. Decision-Case Mix Model for Analyzing Variation in Cesarean Rates. Med Decis Making 2001;21:170-179.

Evans, R.G. Strained mercy: the economics of Canadian health care. Toronto : Butterworths, 1984.

Evans, R.G. The Dog in the Night-time: Medical Practice Variations and Health Policy. In: Andersen, T.F., Mooney, G. (eds). The Challenges of Medical Practice Variations. London: Macmillan, 1990.

Feldman, R., Dowd, B. A new estimate of the welfare loss of excess health insurance. American Economic Review 1991;81:297-301.

Feldman, R., Dowd, B. What does the demand curve for medical care measure? Journal of Health Economics 1993;12:193-200.

Folland, S., Stano, M. Sources of small area variations in the use of medical care. Journal of Health Economics 8 (1989):85-107.

Folland, S., Stano, M. Small Area Variations: a critical review of propositions, method and evidence. Medical Care Review 1990;47:419-65.

Folland, S., Goodman, A.C., Stano, M. The Economics of Health Care. New York : Macmillan, 1993.

Gafni, A., Charles, C., Whelan, T. The physician-patient encounter: The physician as a perfect agent for the patient versus the informed treatment decision-making model. Soc. Sci. Med. 1998;47:347-354.

Gatsonis, C., Normand, S-L., Liu, C., Morris, C. Geographic Variation of Procedure Utilization : a hierarchical model approach. Med Care 1993;31:YS54-YS59, Supplement.

Goodpastor, W.A., Montoya, I.D. Motivating physician behaviour change: social influence versus financial contingencies. International Journal of Health Care Quality Assurance 1996;9:4-9.

Greco, P.J., Eisenberg, J.M. Changing physicians´practice. N Engl J Med 1993;329:1271-1274.

Gruber, J, Kim, J., Mayzlin, D. Physician fees and procedure intensity: the case of cesarean delivery. Journal of Health Economics 1997;18:473-490.

Ham, C. Health Care Variations: assessing the evidence. Research report number 2. London: The King´s Fund Institute, 1988.

Hart, O., Holmström, B. The theory of contracts. In. Bewley, T.B. (ed) - Advances in Economic Theory. Fifth World Congress. Cambridge : Cambridge University Press, 1987.

Johannesson, M. Theory and methods of Economic Evaluation of Health Care. Dordrecht: Kluwer Academic Publishers, 1996.

Johansson, P-O. Evaluating Health Risks: an economic approach. Cambridge: Cambridge University Press, 1995.

Keeler, E.B. A model of demand for effective care. Journal of Health Economics 1995;14:231-238.

Kvinnoklinikerna 1983-1988 : jämförande studie av resurstillgång och resursutnyttjande. Stockholm : Spri, 1991. (Spri rapport 305). (In Swedish).

Labelle, R., Stoddart, G., Rice, T. A re-examination of the meaning and importance of supplier-induced demand. Journal of Health Economics 1994;13:347-368.

Lomas, J., Anderson, G.M., Karin, D.P., et al. "Do practice guidelines guide practice? The effect of a Consensus Statement on the practice of physicians". New Engl J Med 1989;321:1306-11.

Manning, W.G., et al. Health insurance and the demand for medical care: evidence from a randomized experiment. American Economic Review 1987;77:251-277.

Manning, W.G., Marquis, M.S. Health insurance: tradeoffs revisited. Response. Journal of Health Economics 2001;20:289-293.

McClellan, M. The uncertain demand for medical care : a comment on Emmett Keeler. Journal of Health Economics 1995;14:239-242.

McPherson, K. Why Do Variations Occur? In: Andersen, T.F., Mooney, G. (eds). The Challenge of Medical Practice Variations. London : Macmillan, 1990.

Mooney, G., Ryan, M. Agency in health care : getting beyond the first principles. Journal of Health Economics 1993;12:125-135.

Mooney, G. Key Issues in Health Economics. New York: Harvester Wheatsheaf, 1994.

Mulley, A.G. Medical Decision Making and Practice Variation. In: Andersen, T.F., Mooney, G. (eds). The Challenges of Medical Practice Variations. London: Macmillan, 1990.

Nathorst-Böös, J., Munck, I., Eckerlund, I, Ekfeldt-Sandberg, C. An evaluation of the QSP and the QPP: two methods for measuring patient satisfaction. International Journal for Quality in Health Care 2001;13:257-264.

Nyman, J.A. et al. Changing Physician Behaviour: Does Medical Review of Part B Medicare Claims Make a Difference? Inquiry 1990(27):127-137.

Nyman, J.A. The economics of moral hazard revisited. Journal of Health Economics 1999;18:811-824.

Nyman, J.A. The income transfer effect, the access value of insurance and the Rand health insurance experiment. Response. Journal of Health Economics 2001;20:295-298.

Pauly, M.V. What is unnecessary surgery? Milbank Memorial Fund Quarterly/ Health and Society 1979;57:95117.

Peele, P.B. Evaluating welfare losses in the health care market. Journal of Health Economics 1993;12:205-208.

Phelps, C.E. Health Economics. New York : HarperCollins, 1992.

Phelps, C.E., Parente, S.T. Priority Setting for Medical Technology and Medical Practice Assessment. Medical Care 1990;28(8):703-723.

Phelps, C.E., Mooney,C. Correction and Update on 'Priority Setting in Medical Technology Assessment'. Med Care 1992;30:744-751.

Phelps, C.E. Diffusion of Information in Medical Care. Journal of Economic Perspectives 1992;6:23-42.

Phelps, C.E., Mooney, C. Variations in Medical Practice Use : Causes and Consequences. In: Arnould, R.J., Rich; R.F., White; W.D. (eds). Competitive Approaches to Health Care Reform. The Urban Institute Press, Washington D.C., 1993.

Phelps, C.E. Welfare loss from variations : further considerations. Journal of Health Economics 1995;14:253-260.

Rochaix, L. The physician as a perfect agent: a comment. Soc. Sci. Med. 1998;47:355-356.

Reinhardt, U.E. The Economist's Model of Physician Behavior. JAMA 1999;281:462-465.

Rice, T. An alternative framework for evaluating welfare losses in the health care market. Journal of Health Economics 1992;11:85-92.

Rice, T. Demand curves, economists, and desert islands : a response to Feldman and Dowd. Journal of Health Economics 1993;12:201-204.

Rice, T. A model is only as good as its assumptions : a reply to Peele. Journal of Health Economics 1993;12:209-211.

Roos, N.P., Roos, L.L. Small Area Variations, Practice Style and Quality of Care. In: Evans, R.G., Barer, M.L., Marmor, T.R. (eds). Why are some people healthy and others not? New York : Aldine de Gruyter, 1994.

Ryan, M., Mooney, G. Medical Practice Variations - Where now? A paper for debate. HERU. University of Aberdeen. Discussion paper 06/91.

Scott, A. Agency, Incentives and Doctor Behaviour : what do we want doctors to do and how can we get them to do it? Paper presented to the inaugural conference of iHEA, Vancouver, May, 1996. HERU. University of Aberdeen.

Tussing, A.D., Wojtowycz, M.A. Malpractice, Defensive Medicine, and Obstetrical Behavior. Med Care 1997;35:172-191.

Vick, S., Scott, A. Agency in health care. Examining patients'preferences for attributes of the doctor-patient relationship. Journal of Health Economics 1998;17:587-605.

Wennberg, J.E., Gittlesohn, A.M. Small Area Variations in Health Care Delivery. Science 1973;183:1102-1108.

Wennberg, J.E. et al. Changes in Tonsillectomy Rates Associated with Feedback and Review. Pediatrics 1977;59:821-826.

Wennberg, J.E., Barnes, B.A., Zubkoff, M. Professional uncertainty and the problem of supplier-induced demand. Soc Sci Med 1982;16:811-24.

Wennberg, J.E. Dealing with Medical Practice Variations: a proposal for action. In: Health Affairs, Summer 1984;3:6-32.

Wennberg, J.E. The paradox of appropriate care. JAMA 1987;258:2568.

Wennberg, J.E. On the need for outcomes research and the prospects for the evaluative clinical sciences. In: Andersen, T.F., Mooney, G. (eds). The Challenge of Medical Practice Variations. London : Macmillan, 1990.

Wennberg, J.E. Understanding Geographic variation in Health Care delivery, New Engl J Med 1999;340:52-53.

Westert, G.P., Nieboer, A.P., Groenewegen, P.P. Variation in duration of hospital stay between hospitals and between doctors within hospitals. Soc. Sci. med. 1993;37:833-839.

Zweifel, P., Breyer, F. Health Economics. Oxford : Oxford University Press, 1997.

# GLOSSARY

| | |
|---|---|
| **ablatio** | removal (of the placenta) |
| **asphyxia** | failure to breathe (the newborn) |
| **cataract** | condition where the lens of the eye gradually becomes hard and opaque |
| **cesarean section** | surgical operation to deliver a baby by cutting through the abdominal wall into the uterus |
| **disproportion** | condition where the pelvic opening of the mother is not large enough for the head of the fetus |
| **dystocia** | difficult childbirth caused by abnormality (fetus or mother), or malpresentation of the fetus |
| **eclampsia** | serious condition of pregnant woman at the end of pregnancy, where the patient has convulsions and high blood pressure and may go into a coma, caused by toxaemia of pregnancy |
| **fetal distress** | condition of newborn babies where the lungs do not function properly |
| **fetus** | unborn baby |
| **malpresentation** | abnormal presentation of the fetus in the womb |
| **pelvis** | group of bones and cartilage which form a ring and connect the thigh bones to the spine |
| **perinatal** | referring to the period before and after childbirth |
| **placenta** | tissue which grows inside the uterus during pregnancy and links the baby to the mother |
| **placenta praevia** | condition where the fertilized egg becomes implanted in the lower part of the uterus, which means that the placenta may become detached during childbirth and cause brain damage to the baby |
| **pre-eclampsia** | condition of pregnant woman towards the end of the pregnancy, which may lead to eclampsia |
| **strabismus** | squint; condition where the eyes focus on different points |
| **womb, uterus** | hollow organ in the woman's pelvic cavity in which a fertilized ovum (egg) develops into a fetus |

# Essays I-V

International Journal of Technology Assessment in Health Care, 14:4 (1998), 774–787.


I

# ECONOMETRIC ANALYSIS OF VARIATION IN CESAREAN SECTION RATES

## *A Cross-sectional Study of 59 Obstetrical Departments in Sweden*

**Ingemar Eckerlund**
*Stockholm School of Economics and*
*The National Board of Health and Welfare*

**Ulf-G Gerdtham**
*Stockholm School of Economics*

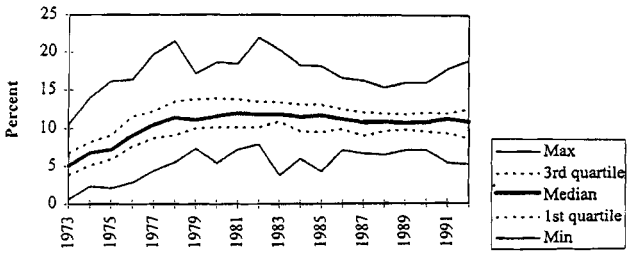
Abstract

The objective of this study was to explain the variation in cesarean section rates among hospitals (obstetrical departments) in Sweden, and to discuss its potential economic consequences. Using data from The Swedish Medical Birth Registry 1991, we made a cross-sectional study of the cesarean section rate at the departmental level. We identified some 20 determinants, demand-related as well as supply-related. A general model including all these regressors was specified. After reducing this model, we were able to explain about one-quarter of the variation. We conclude that the large variation in cesarean section rates indicates inefficiency, due mainly to overutilization, but perhaps also underutilization. It is difficult to calculate the economic consequences or the welfare loss to society. We estimated an additional cost for unnecessary cesarean sections of 13–16 million Swedish crowns (SEK) per year.

**Keywords:** Cesarean Section, Practice Patterns, Economics, Regression Analysis

Deliveries may be performed in one of two ways—vaginally (if necessary, aided by forceps or vacuum extraction) or abdominally by cesarean section. Before 1950, cesarean section was an uncommon procedure in Sweden, accounting for fewer than 1% of deliveries. At that time, a cesarean was a risky intervention, especially for the mother. Even in the late 1950s, the average cesarean section rate was only 2%, ranging from 0.5–4.5% among obstetrical departments.

Since then, however, cesarean sections have become more common in Sweden, as they have in most other comparable countries. International comparisons of cesarean section rates indicate a general increase since the 1970s, as well as large




Source: Medical Birth Registry. National Board of Health and Welfare.

**Figure 1.** Cesarean section rates in Sweden, 1973–92.

variations among (and within) different countries (13). In the United States, permanently among the high-rate countries, for example, the cesarean section rate increased from 5.5% in 1970 to 20.3% in 1983 and 23.5% in 1991 (33). Even in low-rate countries, such as Czechoslovakia, the cesarean section rate has increased rapidly, from 2.3% in 1970 to 6.0% in 1983 and 7.7% in 1988 (37).

In Sweden, the average cesarean section rate more than doubled during the 1970s and amounted to a maximum of 12.3% in 1983. Figure 1 shows the range of variation, and the quartiles, in cesarean section rates among hospitals (obstetrical departments) during the period 1973–92. Obviously, the variation has been considerable and persistent, despite showing a downward trend. The coefficient of variation (COV) decreased from 0.39 in 1973 to 0.24 in 1983 and 0.19 in 1991.

The reasons for this development have been thoroughly dicussed in medical journals (2;26;27;28). Expanded indications, laws, and rules concerning medical liability, and the development of new medical technology are some of the factors believed to explain the rapid rise in the 1970s. Physicians' practice style and personal attitudes toward cesarean section are also supposed to be important explanatory factors (40;41). It should be noted that the decrease during the 1980s has been accompanied by a continuous improvement (decrease) in such outcome measures as perinatal mortality and prevalence of asphyxia (26).

This study investigates the determinants of the variation in cesarean section rate among obstetrical departments in Sweden and discusses the potential economic consequences. We first briefly review the literature. After a theoretical discussion about what governs the choice of mode of delivery, we present a cross-sectional study of cesarean section rates in 1991. The paper concludes with ideas for future research.

## PREVIOUS STUDIES

### Causes

Lomas and Enkin (22) divide the factors explaining the choice of mode of delivery into three categories: obstetrical indications, nonmedical patient-related determinants, and nonmedical professional-related determinants. The major obstetrical indications for cesarean section are: previous cesarean, breech presentation, dystocia, and fetal distress. The nonmedical patient-related determinants are: the woman's socioeconomic status, influence of malpractice litigation, and women's

expectations. Nonmedical professional-related determinants are: financial incentives, convenience, professional discipline, and availability of technology. The authors conclude that the cesarean section rate probably is affected as much by nonmedical factors as by medical factors (36).

Broadhead and James (5) classify the reasons for differences in the utilization of cesarean section among different countries as medical and nonmedical. The medical reasons mentioned are: previous cesarean section, breech presentation, fetal distress, cephalopelvic disproportion, dystocia, low birth weight, and preterm delivery. The nonmedical reasons are: maternal age, socioeconomic status, cultural factors, hospital resources and status, geographical location, private practice, litigation, and physician factors.

McCloskey et al. (24) found that maternal age is a dominant factor independent of clinical risk and type of organization. Zahniser et al. (43) found that women over 35 years had a 30% higher probability of delivering by cesarean section. According to Adashek et al. (1), maternal age, birth weight, need for oxytocin, and spinal anesthesia each showed a positive correlation with cesarean section rate.

Anderson and Lomas (2) compared cesarean section rates by indication at university hospitals and community hospitals. They found that the variation was as large within as between the two types of hospitals. Some studies demonstrate the importance of "the physician factor" (38). Tussing and Wojtowycz (39) found that female doctors usually had a somewhat lower cesarean section rate, but had a higher rate when dystocia was diagnosed. In a study of deliveries performed by 11 physicians in a single community hospital in Detroit, Goyert et al. (17) found that only nulliparity had a greater influence than the identity of the physician on the rate of cesarean section.

Gardner (15) divides the economic factors influencing use of cesarean section into physician-level economic incentives and hospital-level economic incentives. The first group includes reimbursement differential effects and time-minimizing incentives. Even if these may be most pronounced in fee-for-service systems, other economic incentives that favor cesarean section, such as physician convenience, are relevant also in capitated or salary-based systems (35). Examples of hospital-level economic incentives are occupancy rate and the degree of close linkage in the hospital–physician relationship.

Brown (6) found that time-related variables related to the consumption of leisure by physicians, including time of day and day of week, as well as interactions between them, are significant predictors of both total and unplanned cesarean sections.

## Consequences

Cesarean sections have no doubt contributed to the improvements in perinatal morbidity and mortality in this century. However, factors other than increased cesarean section rates are also responsible for this development. Due to improved medical technology, such as anesthetic techniques, pharmacotherapy, etc., maternal morbidity and mortality rates have decreased significantly in abdominal deliveries also. Consequently, the indications for cesarean section have been widened in recent decades, in the best interest of all concerned, and certainly with better outcome for the newborns. However, abdominal delivery is still associated with significantly higher morbidity and mortality for the mother than vaginal delivery. Even if a cesarean section today is regarded as, by and large, a riskless routine procedure, the intervention may cause great discomfort for the woman as well as complications

during and after the operation. The frequency of maternal mortality and complications, such as infertility, is higher mainly for women undergoing acute cesarean section. Furthermore, infants born by cesarean section are at higher risk of developing respiratory distress syndrome (25).

Obviously, variations in cesarean section rates may have medical as well as economic consequences. Regarding economic consequences, it should be noted that a delivery by cesarean section is more resource-demanding than a normal vaginal delivery. The direct costs differ, partly due to difference in the intensity of care and partly due to difference in length of stay. Also, the indirect costs, e.g., in terms of convalescence, are higher after a cesarean section as compared to a vaginal delivery. Overall, maternal morbidity from cesarean section is associated both with higher direct costs of care and with such indirect costs as production loss, delayed recovery, and impaired maternal–infant bonding (11;12;15;31).

Based on British experience, Clark et al. (7) have made some cost calculations for different modes of delivery, showing that the hospital cost of a cesarean section without complications is three times that of a normal vaginal delivery. According to U.S. diagnosis-related groups (DRG) payments, the additional cost for a cesarean section is 85% over vaginal delivery. Based on a study by Health Insurance Association of America, Keeler and Brodie (20) found that a cesarean section is 66% more expensive than a normal delivery. Corresponding Swedish DRG data (1993) show that the cost of a cesarean section is SEK 36,000 (with complications, DRG 370) and SEK 26,500 (without complications, DRG 371), compared to the cost of a vaginal delivery, which was SEK 12,500 (with complications, DRG 372) and SEK 10,000 (without complications, DRG 373).

In the light of these figures it seems quite obvious that a higher cesarean section rate implies higher costs for obstetric care. According to Gardner (15), the annual savings associated with each percentage point reduction in the cesarean section rate in the United States would amount to more than $68 million (1987 dollars). This estimate does not include the additional savings that would result from lower physician fees and avoided cesarean section–related morbidity and mortality.

## THEORETICAL FRAMEWORK

To further understand medical practice variations, it is important to understand what governs medical decision making, i.e., the physicians' practice behavior. Various theories have been launched in the literature concerning the major determinants for medical decision making (9;10;14). Theoretically, the medical decision can be described as an optimization problem under uncertainty. Physicians/departments are assumed to maximize a utility function, including their own utility as well as patients' utility and the social utility. The utility maximization is constrained by a variety of restrictions—medical, organizational, technological, economic, etc. Using terminology from contract theory, the decision-making physician/department can be regarded as a "double agent" for several principals with at least partly conflicting interests (3).

Folland and Stano (14) developed an econometric model of physicians' resource utilization, including uncertainty and practice style as explanatory factors behind existing variations in utilization of medical and surgical procedures. Practice style is regarded as reflecting the physician's beliefs concerning the health production function, i.e., the correlation between the care provided and its benefit to the patient. The fact that physicians' beliefs differ is assumed to be due to professional

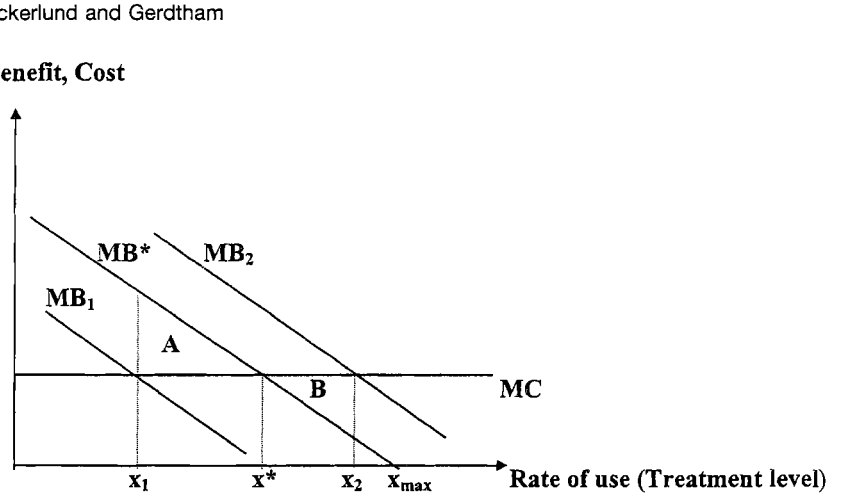



**Figure 2.** Marginal benefit and marginal cost as functions of use rate.

uncertainty. Based on empirical tests of the model, the authors conclude that practice style does not seem to explain variations in resource utilization at an aggregated level, but probably would do so for single procedures.

The "professional uncertainty" hypothesis has been confirmed by a number of studies, demonstrating that variation, such as in hospital admission rates, tends to be smaller when there is professional consensus on what constitutes best practice (29;34;40). Different beliefs about the production function give rise to different marginal benefit (MB) curves (Figure 2). Assuming that MB* is derived from the "true" production function, the optimal treatment level is x*, where marginal benefit equals marginal cost (MC). The treatment levels given by the intersection between $MB_1$ ($MB_2$) and the MC curve, $x_1$ ($x_2$), thus indicates underutilization (overutilization). The welfare loss in each case is represented by the triangles A and B, respectively. If the patient has complete information and pays for the care out of pocket, then she will demand x*. If, instead, the patient is fully insured (which, in principle, is the case in Sweden), she will maximize well-being by demanding $x_{max}$. A physician acting as a perfect agent for the patient would also provide or recommend this quantity ($x_{max}$) of care. However, in the Swedish setting the physician should probably be regarded as a "double agent" for several principals (the patient, the hospital/department, the County Council, and the community). A compromise solution would be somewhere between x* and $x_{max}$, which means that there is an incentive toward overutilization.

Moreover, there are other incentives working in the same direction. From the physician's perspective, two kinds of mistakes (wrong decisions) are possible: to perform a certain procedure when it is "inefficient" (to the right of x* in Figure 2), and not to perform it when it would be "efficient" (to the left of or exactly at x* in Figure 2). Both kinds of mistakes may have medical as well as economic consequences. However, the consequences are not the same, and attempts to reduce the first kind of mistake may increase the probability of the second, and vice versa. The incentives to avoid the second kind of mistake, and thereby reduce the probability of malpractice litigation, seem to be much stronger than the incentives to avoid the first kind of mistake, which means that there is a built-in bias toward overutilization.

Applying this simple model to the case of cesarean sections in Sweden, we assume that the existing variation in cesarean section rates reflects differences not only in demand or patients' needs, but also in supply-related factors, including physicians' beliefs about the health benefit of cesarean sections. In the next section, we will specify and estimate an econometric model to explain the variation in cesarean section rates among obstetrical departments in Sweden. We first should provide some background information on the Swedish health care system. The hospitals in Sweden are publicly owned and operated, traditionally within fixed budgets. The physicians are employed by the hospital and salaried. There are specially appointed chief physicians responsible for all health care in their departments. Only a few obstetricians at each obstetrical department are entitled to decide whether to perform a cesarean section (8). This makes it reasonable to assume a consistent practice style within the single department, i.e., an insignificant intradepartmental variation among individual doctors. This in turn means that the relevant study unit is the clinical department rather than the single physician.[1]

We assume that the obstetrician/department—in consultation with the patient—chooses the mode of delivery. Decision making is primarily guided by *demand-related factors*, such as age, morbidity, expectations, etc. The physician/department may, however, also want to fulfill personal interests and act as an agent for the employer, the hospital, and the County Council, who may be assumed to demand the cheapest medically defensible method to be used. This implies that resource capacity and factors related to clinical judgment and practice style—what we call *supply-related factors*—also influence the decisions.

## CROSS-SECTIONAL STUDY, 1991

### Material and methods

***Data.*** Using data from The Swedish Medical Birth Registry 1991, a special database was compiled, containing the following information on all newborns in Sweden that year: cesarean section, maternal age, parity (order of delivery), stillbirth, perinatal mortality (stillbirth or dead within first week), mother and father living together, number of newborns, prevalence of asphyxia (Apgar score <7 at 5 minutes), hospital (five-digit code), and indication for cesarean section (placenta previa = 1, ablatio = 2, disproportion = 3, pre-eclampsia = 4, dystocia = 5, fetal distress = 6, twins or more = 7, malpresentation = 8, psychosocial indication = 9, and other indication = 10).

In 1991, 13,895 newborns were delivered by cesarean section in Sweden, corresponding to 11.2% of all (123,650) newborns. In order to make relevant comparisons on the department level, we excluded units with fewer than 600 deliveries per year. Our material thus includes data from 59 obstetrical departments, covering 97% of all newborns. The average cesarean section rate was 11.2%, ranging from 5.6% to 18.0% across departments.

On average, the university hospitals and most well-equipped general hospitals had a slightly higher cesarean section rate (11.3%) than other hospitals (10.6%). The variation within the first group of hospitals was larger (5.6–18.0%; COV = 0.20) than among the second group (8.7–13.7%; COV = 0.16). The variation among the 26 counties ranged from 7.7% to 13.3%, and among the six health care regions, from 10.2% to 12.3%.

**Table 1.** Incidence of Different Indications for Cesarean Section

| Indication | Number of cases | Cesareans (%) | % of all cesareans |
|---|---|---|---|
| Placenta praevia | 299 | 84 | 2 |
| Ablatio | 657 | 77 | 4 |
| Disproportion | 4,038 | 88 | 25 |
| Pre-eclampsia | 2,957 | 33 | 7 |
| Dystocia | 8,792 | 20 | 10 |
| Fetal distress | 5,890 | 40 | 15 |
| Twins | 2,074 | 33 | 5 |
| Malpresentation | 15,258 | 22 | 25 |
| Psychosocial | 999 | 53 | 4 |
| Other indication | 78,351 | 1 | 3 |
| | 119,315 | 11.2 | 100 |

Table 1 shows the frequency of different indications for cesarean section, the percentage of cesarean section per obstetrical indication, and the distribution of the cesarean sections over different indications. The 10 classes of indications more or less frequently resulted in cesarean section. The highest frequency was disproportion. The most common indications, given that cesarean section was performed, were disproportion (25%), malpresentation (25%), and fetal distress (15%).

***Measurement and Model Specification.*** Based on the literature review, our theoretical discussion, and the database described above, we defined a dependent variable (cesarean section rate) and a number of independent variables. These variables were used to specify and estimate an econometric model, aiming at explaining the variation in cesarean section rate among obstetrical departments. The variables are shown in Table 2. The cesarean section rate increases with maternal age. It is considerably higher for mothers over 35 years. Even if the risk is a continuously increasing function of age, we defined the variable *age* as the percentage of mothers over 35 years, which also happens to be the most common measure in previous studies. Cesarean section is more frequent among nullipara (mothers having their first child). *Parity*, defined as the percentage of nullipara, is thus supposed to be positively correlated with cesarean section rate. Further, cesarean section is more common in multiple births (twins or more), 40.1%, compared to 10.4% in single births. This is captured in the variable *twins*. The variables *indication 1–indication 9* are defined as percentage of cases with these indications, respectively. The indications are based on obstetrical diagnoses and are mutually exclusive on the individual level. The last group, *other indication*, is a residual and is not included in the analysis.

Since some high-risk cases are referred from smaller to bigger and technically better equipped hospitals, it seems natural that these latter hospitals have a higher rate of cesarean sections. The dummy variable *refhosp* indicates whether a hospital has continuous access to an obstetrician, as well as to a pediatrician and an anesthesiologist.

Workload and availability of beds may have an influence on the propensity to perform a cesarean section. Number of deliveries per obstetrician, *delphys*, and the occupancy rate at maternity wards, *occupancy*, are intended to measure these factors. Since the physicians also work at gynecological wards and clinics, we included two more workload variables, the number of admissions at gynecological wards per doctor, *gadmphys*, and the number of patient visits at the gynecological

**Table 2.** Variables: Average Values and Coefficients of Variation, 1991

| Variables | Description | Average | Coefficients of variation |
|---|---|---|---|
| *Dependent variable* | | | |
| Cesarean section rate | Percentage of cesarean sections | 11.2 | 0.19 |
| *Independent variables* | | | |
| | Percentage of: | | |
| Age | mothers > 35 yr | 11.3 | 0.19 |
| Parity | nullipara | 40.2 | 0.08 |
| Twins | twin (or more) births | 2.3 | 0.42 |
| | Percentage with: | | |
| Indication 1 | Indication 1 (placenta previa) | 0.3 | 0.64 |
| Indication 2 | Indication 2 (ablatio) | 0.5 | 0.58 |
| Indication 3 | Indication 3 (disproportion) | 3.3 | 0.45 |
| Indication 4 | Indication 4 (pre-eclampsia) | 2.4 | 0.40 |
| Indication 5 | Indication 5 (dystocia) | 6.9 | 0.77 |
| Indication 6 | Indication 6 (fetal distress) | 4.7 | 0.40 |
| Indication 7 | Indication 7 (twins or more) | 1.6 | 0.43 |
| Indication 8 | Indication 8 (malpresentation) | 13.7 | 0.57 |
| Indication 9 | Indication 9 (psychosocial) | 0.7 | 1.07 |
| Refhosp | Well-equipped hospital? (yes = 1, no = 0) | 0.7 | 0.06 |
| Delphys | Deliveries per physician | 143.7 | 0.27 |
| Gadmphys | Gyn admissions per physician | 109.7 | 0.16 |
| Gvisphys | Gyn visits per physician | 886.8 | 0.41 |
| Occupancy | Occupancy rate at maternity wards; % | 79.5 | 0.21 |
| Management | Senior obstetrician? (yes = 1, no = 0) | 0.8 | 0.51 |
| Education | Continuous follow-up? (yes = 1, no = 0) | 0.8 | 0.54 |
| Malpract | Any complaint in 1990? (yes = 1, no = 0) | 0.2 | 1.79 |

clinics per doctor, *gvisphys*. These variables are expected to be negatively correlated with the cesarean section rate.

We also considered variables more closely related to the physician practice style. The two dummy variables, *management* and *education*, measure whether there is a senior obstetrician with special responsibility for the delivery unit, and whether there is a systematic follow-up of all cesarean sections performed within the unit as part of the internal education. Finally, we included a dummy variable, *malpract*, indicating whether there were any complaints (concerning cesarean sections) against the department submitted to the Responsibility Board for Health and Medical Care in 1990, the previous year. The idea was that such a complaint may, in the short run, bring about a more liberal attitude toward performing cesarean sections. Some other variables were tested, such as the percentage of routine use of electronic fetal monitoring. Problems concerning the definition of "routine" made us exclude that variable from the analysis. However, available (incomplete) data did not support the hypothesis of a significant correlation between this variable and cesarean section rate.

To ensure that predictions remain inside the feasible range (0–100%), the dependent variable (cesarean section rate) was transformed, by taking the natural logarithm of the rate divided by its complement, into a "logit":

$$\textbf{Insectio} = \ln(\text{cesarean section rate}/(100 - \text{cesarean section rate})) = \mathbf{b}'\mathbf{x} + \mathbf{u}$$

where **b**, **x** and **u** denote the regression coefficients, the explanatory variables, and the disturbance term, respectively. We used aggregate data from the 59 obstetrical departments to estimate the logit model. Since the disturbance term in the logit model is heteroscedastic, with variance equal to $1/[N_i\ psect_i\ (1 - psect_i)]$, we have to use weighted least-squares (WLS) regression. The inverse of the variance is used as weights (18).[2]

We started from an unrestricted equation, including all independent variables enumerated in Table 2, which was then reduced by successive elimination of variables not significantly correlated (on the basis of individual *t* statistics) with the dependent variable. Tests for misspecification include tests for non-normality of the residuals by the Jarque-Bera's skewness-kurtosis test (19), heteroscedasticity by the Breusch-Pagan LM test (4), and functional misspecification by the Ramsey RESET test (16;21).

## Results

The results are summarized in Table 3. The final reduced model can be written as shown below the table. The cesarean section rate, i.e., the log of the odds, increases with the percentage of mothers over 35 years and with the percentage of cases with indication 1 (placenta previa), but decreases with the existence of systematic follow-up of performed cesareans. The model explains 27% of the variation.[3]

The hypothesis that all coefficients for variables excluded in the reduced model are equal to zero (slope restrictions) was tested using a joint F test, and could not be rejected at the 5% level. The Jarque-Bera test (J-B), applied to test for normality in residuals, showed that the null hypothesis could not be rejected at the 5% level. From the Breusch-Pagan test (B-P), it appeared that none of the models showed heteroscedastic residual variance. The RESET test did not indicate functional form misspecification for any of the models.

## Discussion

The purpose of our study was to explain the interdepartmental variation in cesarean section rates, and to discuss its potential economic consequences. The final model includes a constant term and three independent variables, all with expected signs. Thus, our findings suggest that systematic follow-up is well suited to help control the cesarean section rate. The model explains only about one-quarter of the observed variation. Many of the independent variables considered could be defined differently. We tried a number of alternative definitions, e.g., of the age variable, without any significant effect on the explanatory value. We also tried to cluster some of the indications, again without much effect.

Even within different indication groups, we found large variations in cesarean section rates among departments. This indicates that the decision on mode of delivery to a large extent is governed by what we call *practice style*, i.e., the attitudes or beliefs of the responsible obstetricians. The variation in cesarean section rates within the various indication groups is probably due not only to differences in probability of cesarean section, given a certain indication (*treatment practice style*), but also to differences in propensity to set a certain diagnosis (*diagnosis practice style*). Furthermore, it cannot be excluded that the variation is partly due to vague criteria for different diagnoses or faulty registration routines. A high cesarean section rate, given a certain indication, may be partly due to a low propensity to

**Table 3.** Multivariate Regression: Dependent Variable, Insectio[a]

| | General model | | Reduced model | |
|---|---|---|---|---|
| Parameter/variable | Coefficient | *t*-value | Coefficient | *t*-value |
| Constant | −2.840[b] | −6.364 | −2.413[b] | −20.586 |
| Age | 0.028[c] | 1.934 | 0.028[b] | 3.035 |
| Parity | 0.011 | 0.923 | | |
| Twins | 0.016 | 0.139 | | |
| Indication 1 | 0.416[d] | 2.142 | 0.408[d] | 2.399 |
| Indication 2 | 0.127 | 1.282 | | |
| Indication 3 | 0.009 | 0.362 | | |
| Indication 4 | 0.060 | 1.497 | | |
| Indication 5 | −0.007 | −1.078 | | |
| Indication 6 | −0.022 | −1.191 | | |
| Indication 7 | 0.043 | 0.302 | | |
| Indication 8 | 0.002 | 0.365 | | |
| Indication 9 | 0.021 | 0.543 | | |
| Refhosp | −0.176[c] | −1.887 | | |
| Delphys | −6.821E-04 | −0.663 | | |
| Gadmphys | 7.215E-04 | 1.073 | | |
| Gvisphys | −1.636E-04 | −1.386 | | |
| Occupancy | 0.002 | 0.718 | | |
| Management | −0.073 | −0.677 | | |
| Education | −0.161[d] | −2.291 | −0.104[c] | −1.831 |
| Malpract | 0.116 | 0.168 | | |
| d.f. | 38 | | 55 | |
| $R^2$ ($R^2$ adj) | 0.466 (0.185) | | 0.269 (0.230) | |
| SSE | 230.994 | | 316.157 | |
| Slope restr | — | | F(17;38) = 0.824 | |
| J-B (d.f.) | $\chi^2(2)$ = 2.454 | | $\chi^2(2)$ = 2.762 | |
| B-P (d.f.) | $\chi^2(20)$ = 12.371 | | $\chi^2(3)$ = 3.050 | |
| RESET (d.f.) | F(3;35) = 0.007 | | F(3;52) = 0.327 | |

[a] Final reduced model:
**insectio** = −2.41 + 0.03 **age** + 0.41 **indication** 1 − 0.10 **education**
(−20.59) (3.04) (2.40) (−1.83) ($R^2$ = 0.27)
[b] Significant at the 1% level.
[c] Significant at the 5% level.
[d] Significant at the 10% level.

set (and/or lacking carefulness in the registration of) diagnoses, leading to this indication. And vice versa, a low cesarean section rate may be due to a generous attitude to setting (and/or carefulness to register) the actual diagnoses. In summary, we have reason to believe that much of the unexplained variation in cesarean section rates has to do with differences in practice patterns. However, these are difficult to identify and even more difficult to quantify and model.

Differences in cesarean section rates may have medical as well as economic consequences. Most interesting from an economic perspective is the extent to which the variation indicates productivity or efficiency problems. Our study focuses mainly on the potential efficiency problems related to the overutilization of cesarean sections. An evidently necessary cesarean section that is not performed may, of course, create an even more serious problem from an efficiency perspective. However, this situation is probably quite rare.

Since we were unable to fully identify the causes of variation, we have a limited opportunity to evaluate the consequences of variation. In the individual case, an

unnecessary cesarean section is an expression of inefficient resource utilization. The practitioner has intervened beyond the indications, i.e., used the method where the marginal value to the patient is exceeded by the marginal cost, including the increased risk. Unnecessary costs are incurred: *direct costs* of the procedure itself and patient care, which are higher than those of the alternative, vaginal delivery, and *indirect costs* related to the increased risk for complications during and following the surgery, which in turn generate costs.

However, it is difficult to estimate the *opportunity cost* (i.e., the value of the resources in the best alternative use), which may vary over time depending on the workload of the clinical departments. It is also equally difficult to determine the number of unnecessary cesarean sections. No universal agreement exists in applying the indications for cesarean section. Furthermore, the tendency to meet patients' preferences for cesarean section, generated by the fear or anxiety of childbirth, may vary among the clinical departments. Is a cesarean section that is not strictly motivated by medical indications (not "appropriate"), but which meets a patient's preference, "unnecessary"? Opinions are also divided concerning the value of cesarean section for prolonged deliveries where other complications appear.

Lacking anything better, we used the average *additional cost* of a cesarean section as an approximation of the opportunity cost. Using Swedish DRG data from 1993, this cost—defined as the difference in cost between DRG 371, cesarean section without complications, and DRG 372 (373), vaginal delivery with (without) complications—has been estimated at 14,202 SEK (95% confidence interval, 13,568–14,836) and 16,671 SEK (95% confidence interval, 16,124–17,218), respectively. The number of unnecessary cesarean sections is, for reasons mentioned above, difficult to calculate. If we assume that it were possible—without complications for mother or child—to lower the number of cesarean sections at the clinical departments that exceed the average rate (11.3% at the referral hospitals and 10.6% at the others), it would mean about 940 fewer cesarean sections per year (1991). This corresponds to an additional cost for unnecessary cesarean sections of 13–16 million SEK.

We also considered an alternative approach to calculating the economic consequences of variation in cesarean section rates, based upon a method developed by Phelps and Parente (32), and corrected by Phelps and Mooney (30). They have shown that the total welfare loss (W) to society resulting from interarea deviations from the correct[4] rate is equal to

$$\mathbf{W = 0.5\ (total\ expenses\ for\ X)\ (COV)^2/\eta}$$

where COV is the coefficient of variation of inappropiate use (i.e., the standard deviation in unexplained use divided by the average use of X)[5], and $\eta$ is the absolute value of the price elasticity of demand. Phelps and Mooney (32) recommend $\eta$ = 0.15 as a global estimate of demand elasticities, obtained from the RAND Health Insurance Study (HIS) (23). Unfortunately, we have no corresponding estimate of $\eta$ in Sweden. If we, nonetheless, assume $\eta$ = 0.15, we get W = 31.2 million SEK, due to unexplained variation in cesarean section rates. If, instead, we substitute our above estimate of additional cost (13–16 million SEK)[6] for W in the formula, and calculate $\eta$ we get $\eta$ = 0.29–0.36. Since the Swedish patients are in principle fully insured, the "true" $\eta$ is probably closer to 0.15 than to 0.29–0.35. To summarize, the welfare loss due to undesired variation in cesarean section rates among obstetrical departments is at least 13–16 million SEK, and probably double.

## CONCLUDING REMARKS

In this study, we used data from the Swedish Medical Birth Registry, including all deliveries in 1991. Alternatively, we could have made a sample study, using questionnaires or interviews to gather the information needed. Both approaches have their pros and cons. A disadvantage with the method chosen is the difficulty in capturing exactly what governs the decision of whether to carry out a cesarean section.

The choice of method for statistical analysis is governed by which decision level is examined. Most available studies on variations in cesarean section rates seem to focus on the individual physician/patient and use some kind of logistic regression with a dichotomous dependent variable. In our study, however, the focus was on the clinical department level, and consequently we used multiple linear regression analysis.

To find out in more detail which factors govern the choice of mode of delivery, we probably have to carry out a survey study based on a sample of deliveries. In particular, careful studies of a number of individual cases might enable us to learn more about the influence of practice patterns on the variation in cesarean section rates. Finally, it should be noted that variations per se tell us nothing about the potential for greater efficiency. If the distribution as a whole is off target, a small variation may be a greater problem than a large variation where the distribution is generally on target. No particular level is right for all clinical departments. It may be possible to determine what is right after the fact, when we have access to the final results. Some variation is, of course, desirable—assuming it reflects the differences in medical needs and patients' preferences.

### NOTES

[1] This also influenced our choice of statistical method. We considered employing multilevel analysis techniques. However, since we lack data on the physician level and since our focus is on the department level, we chose to use multiple linear regression analysis. Moreover, studies using multilevel statistical techniques have demonstrated that the within hospital variation (in length of hospital stay) between doctors is in most cases statistically insignificant (42).

[2] The weights, $w_i = N_i\ psect_i\ (1 - psect_i)$; $N_i$ is the number of deliveries at department i, and $psect_i$ is the relative frequency of cesarean sections.

[3] We also ran linear probability (LP) models with cesarean section rate (percent) as dependent variable. This yielded the same model, but a slightly higher explanatory value ($R^2 = 0.31$). The LP model has the advantage that the coefficients are easier to interpret, but the disadvantage that the predicted values of the cesarean section rates may turn out to be less than 0 or greater than 100. The Logit model has the disadvantage that the estimated parameters are sensitive to small measurement errors if the cesarean section rates are close to zero or 100.

[4] Phelps and Parente (32) first assume that the average amount of any medical intervention (X) across all regions is correct, and then show that this assumption probably implies an underestimate of the true welfare loss.

[5] This modified $COV^2 = (1 - R^2)\ COV^2$.

[6] This sum refers to overutilization only.

### REFERENCES

1. Adashek, J. A., Peaceman, A. M., & Lopec-Zeno, J. A., Factors contributing to the increased cesarean birth rate in older parturient women. *American Journal of Obstetrics and Gynecology*, 1993, 169, 936–40.

2. Anderson, G. M., & Lomas, J. Determinants of the increasing cesarean birth rate. *New England Journal of Medicine*, 1984, 311, 887–92.
3. Blomqvist, Å. The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics*, 1991, 10, 411–22.
4. Breusch, T., & Pagan, A. A simple test for heteroskedasticity and random coefficient Variation. *Econometrica*, 1979, 47, 1287–94.
5. Broadhead, T. J., & James, D. K. Worldwide utilization of cesarean section. In B. L. Flamm & E. J. Quilligan (eds.), *Cesarean section: Guidelines for appropiate utilization.* New York: Springer-Verlag, 1995, 9–21.
6. Brown, H. S. Physician demand for leisure: Implications for cesarean section rates. *Journal of Health Economics*, 1996, 15, 233–42.
7. Clark, L., Mugford, M., & Paterson, C. How does the mode of delivery affect the cost of maternity care? *British Journal of Obstetrics and Gynaecology*, 1991, 98, 519–23.
8. Eckerlund, I., & Nielsen, T. F. *Quality of care at delivery units: A survey study of obstetrical departments in Sweden.* Unpublished memo, 1995.
9. Eisenberg, J. *Doctors' decisions and the cost of medical care: The reasons for doctors' practice patterns and ways to change them.* Ann Arbor, MI: Health Administration Press Perspectives, 1986.
10. Ellis, R. P., & McGuire, T. G. Provider behaviour under prospective reimbursement. *Journal of Health Economics*, 1986, 5, 129–51.
11. Evans, R. G. The dog in the night: Medical practice variations and health policy. In T. F. Andersen & G. Mooney (eds.), *The challenges of medical practice variations.* London: Macmillan, 1990.
12. Finkler, M. D., & Wirtschafter, D. D. Cost-effectiveness and obstetric services. *Medical Care*, 1991, 29, 951–63.
13. Flamm, B. L., & Quilligan, E. J. (eds). *Cesarean section: Guidelines for appropriate utilization.* New York: Springer-Verlag, 1995.
14. Folland, S., & Stano, M. Sources of small area variations in the use of medical care. *Journal of Health Economics* 1989, 8, 85–107.
15. Gardner, L. B. Economic considerations in cesarean section use. In: B. L. Flamm & E. J. Quilligan (eds.), *Cesarean section: Guidelines for appropiate utilization.* New York: Springer-Verlag, 1995, 173–190.
16. Gerdtham, U.-G. *Essays on international comparisons of health care expenditure.* Ph.D. dissertation. Linköping: Universitetet i Linköping, 1991 (Linköping Studies in Art and Science, No. 66).
17. Goyert, G. L., Bottoms, S. F., Treadwell, M. C., & Nehra, P. C. The physician factor in cesarean birth rates. *New England Journal of Medicine*, 1989, 320, 706–09.
18. Gujarati, D. N. *Basic econometrics*, 2nd ed. New York: McGraw-Hill, 1988.
19. Jarque, C. M., & Bera, A. K. A test for normality of observations and regression residuals. *International Statistical Review*, 1987, 55, 163–72.
20. Keeler, E. B., & Brodie, M. Economic incentives in the choice between vaginal delivery and cesarean section. *The Milbank Quarterly*, 1993, 71, 365–404.
21. Kmenta, J. *Elements of econometrics.* 2nd ed. New York: Macmillan, 1986.
22. Lomas, J., & Enkin, M. Variations in operative delivery. In I. Chalmers, M. Enkin, & M. Keirse (eds.) *Effective care in pregnancy and childbirth*, vol. 2. Oxford: Oxford University Press, 1991.
23. Manning, W. G., Newhouse, J. P., Dvan, N., et al. Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review*, 1987, 77, 251.
24. McCloskey, L., Petitti, D. B., & Hobel, C. J. Variations in the use of cesarean delivery for dystocia: Lessons about the source of care. *Medical Care*, 1992, 30, 126–35.
25. Nielsen, T. F. Kejsarsnitt i modern obstetrik: Vad är vetenskap och beprövad erfarenhet? *Läkartidningen* 1985, 82, 1441–43.
26. Nielsen, T. F., Otterblad Olausson, P., & Ingemarsson, I. The cesarean section rate in Sweden: The end of the rise. *Birth*, 1994, 21, 1.

27. Notzon, F. C., Placek, P. J., & Taffel, S. M. Comparison of national cesarean-section rates. *New England Journal of Medicine*, 1987, 316, 386–89.
28. Notzon, F.C., Cnattingius, S., Bergsjö P., et al. Cesarean section delivery in the 1980's: International comparison by indication. *American Journal of Obstetrics and Gynecology* 1994, 170, 495–504.
29. Phelps, C. E. Diffusion of information in medical care. *Journal of Economic Perspectives*, 1992, 6, 23–42.
30. Phelps, C. E., & Mooney, C. Correction and update on 'Priority Setting in Medical Technology Assessment. *Medical Care*, 1992, 30, 744–51.
31. Phelps, C. E., & Mooney, C. Variations in medical practice use: Causes and consequences. In R. J. Arnould, R. F. Rich, W. D. White (eds.), *Competitive approaches to health care reform.* Washington DC: Urban Institute Press, 1993.
32. Phelps, C. E., & Parente, S. T. Priority setting in medical technology and medical practice assessment. *Medical Care*, 1990, 28, 703–23.
33. *Rates of cesarean delivery—United States*, 1991. *MMWR*, 1993, 42, 285–89.
34. Ryan, M., & Mooney, G. *Research in medical practice variations: Where now? A paper for debate.* Aberdeen. HERU, University of Aberdeen. Discussion paper 06/91.
35. Stafford, R. S. Cesarean section use and source of payment: An analysis of California hospital discharge abstracts. *American Journal of Public Health*, 1990, 80, 313–15.
36. Stafford, R. S. The impact of nonclinical factors on repeat cesarean section. *JAMA*, 1991, 265, 59–63.
37. Stephensen, P. A. *International differences in the use of obstetrical interventions.* Copenhagen: WHO Regional Office for Europe, 1991 (EUR/ICP/MCH 112).
38. Tussing, A. D., & Wojtowycz, M. A. The cesarean decision in New York State, 1986: Economic and noneconomic aspects. *Medical Care*, 1992, 30, 529–40.
39. Tussing, A. D., & Wojtowycz, M. A. The effect of physician characteristics on clinical behavior: Cesarean section in New York state. *Social Science and Medicine*, 1993, 37, 1251–60.
40. Wennberg, J. E. Dealing with medical practice variations: A proposal for action. *Health Affairs*, 1984, 3, 6–32.
41. Wennberg, J. E., Barnes, B.A., & Zubkoff, M. Professional uncertainty and the problem of supplier-induced demand. *Social Science and Medicine*, 1982, 16, 811–24.
42. Westert, G. P., Nieboer, A. P., & Groenewegen, P. P. Variation in duration of hospital stay between hospitals and between doctors within hospitals. *Social Science and Medicine*, 1993, 37, 833–39.
43. Zahniser, S. C., Kendrick, J. S., Franks, A. L., et al. Trends in obstetric operative procedures, 1980 to 1987. *American Journal of Public Health*, 1992, 82, 1340–44.

International Journal of Technology Assessment in Health Care, 15:1 (1999), 123–135.


# ESTIMATING THE EFFECT OF CESAREAN SECTION RATE ON HEALTH OUTCOME

II

## *Evidence from Swedish Hospital Data*

**Ingemar Eckerlund**
*Stockholm School of Economics and The National Board of Health and Welfare*

**Ulf-G. Gerdtham**
*Stockholm School of Economics*

Abstract

This paper tests the null hypothesis of a zero effect of cesarean section rate on health outcome against the alternative of a positive effect. Using data from 59 hospitals in Sweden from 1988–92, we specify two separate linear regression models for health outcome, one with perinatal mortality, and the other with rate of asphyxia, as dependent variable. We estimate the models by single-year cross-section regressions and as pooled data systems. The null hypothesis cannot be rejected, i.e., we do not find any significant positive effect of cesarean section rate on health outcome. Thus, we conclude that an increase in cesarean section rate does not imply lower perinatal mortality or lower rate of asphyxia. This in turn indicates that the minimum cesarean section rate is optimal.

**Keywords:** Cesarean Section, Health Outcome, Regression Analysis, Economic Consequences

In Sweden, as in many other comparable countries, there has been a considerable and persisting variation in cesarean section rates among hospitals (obstetrical departments) (3;7;14). In a cross-sectional study of 59 hospitals in 1991, we identified some 20 determinants of cesarean section rate, demand-related as well as supply-related (5). A general model including all these regressors was specified. After reducing this model, we were able to explain about one-quarter of the variation. Since only a minor part of the variation could be explained by differences in demand- and supply-related variables, and since a cesarean section is more resource-demanding than a vaginal delivery, we concluded that the variation in cesarean section rates indicates inefficiency. What, then, is the optimal rate of cesarean section?

Financial support from the Swedish Council for Social Research and the Swedish Institute for Health Services Development is gratefully acknowledged.

To answer that question, we must estimate the effect of cesarean section rate on health outcome, i.e., estimate the health production function. While there have been numerous studies of the determinants of variation in cesarean section rates, there is little information on the consequences of differing cesarean rates on outcome variables. A critical issue is how to define appropriate outcome measures. The main reason to perform a cesarean section is to improve the health outcome for the newborn,[1] and the traditional measure used to assess outcome in obstetric care is perinatal mortality (15). Perinatal mortality rates have decreased in recent decades while there has been an upward trend in cesarean section rates. It seems logical to apply a cause-and-effect relationship between these phenomena. However, on closer investigation, there is not much to support such a relationship. Available data suggest that the cesarean section rate does not contribute much to explaining the variation in perinatal mortality (2).

International comparisons demonstrate that perinatal mortality has generally improved, regardless of whether the cesarean section rates have increased, decreased, or remained stable over time (16). Several studies have confirmed this finding and failed to show a strong correlation between cesarean section rate and perinatal mortality. Furthermore, it has been shown that it is possible to lower the cesarean section rate on a nationwide basis without increasing risks to newborn infants (15).

Joffe et al. (10) compared groups of obstetric units, categorized according to their cesarean section rates, to see what effect the rate has on fetal and maternal outcomes.[2] They found that for very low-birth-weight-infants, but not for other deliveries, mortality rates were lower in units with higher cesarean section rates.[3] Apgar scores showed no trend, but the onset of respiration after 1 minute was significantly more frequent in units with a cesarean section rate of less than 10%. Increased maternal postnatal blood transfusion was associated with higher cesarean section rates, but no trend was observed for the other maternal variables.

In summary, there is good reason to consider outcome measures other than perinatal mortality, such as a measure of morbidity. Furthermore, to assess the effectiveness of cesarean section with respect to outcome (mortality and morbidity), it is necessary to control for case-mix, or alternatively, to look at outcome for specific indications rather than overall outcome.

This paper tests the null hypothesis of a zero effect of cesarean section rate on health outcome against the alternative of a positive effect. We first briefly describe the data and the variables selected. A description of the estimation methods used are followed by the results. Finally, a discussion of our findings and some concluding remarks on the need for further research are presented.

## DATA AND MEASUREMENT

Using data from the Swedish Medical Birth Registry from 1988–92 (14), we compiled a special database containing yearly data from the 59 hospitals—covering 97% of all newborns—in the abovementioned study (5) on the number of newborns, deliveries, perinatal deaths, cesarean sections, acute cesarean sections, mothers older than 35 years, multiple births (twins or more), cases with asphyxia (Apgar score $< 7$ at 5 minutes),[4] and cases with indications 1–11, respectively (placenta praevia = 1, ablatio = 2, disproportion = 3, pre-eclampsia = 4, dystocia = 5, fetal distress =

**Table 1.** Description of Dependent and Independent Variables, Average Value, and Coefficient of Variation, 1988–91

| Variable | Description | Average value | CV |
|---|---|---|---|
| *Dependent variables* | | | |
| Permort | Percentage: Perinatal deaths | 0.59 | 0.44 |
| Asphyx | With asphyxia | 1.08 | 0.39 |
| *Independent variables* | | | |
| C-section | Percentage: Cesarean sections | 10.74 | 0.20 |
| Age | Mothers > 35 years | 11.13 | 0.18 |
| Twins[a] | Twin (or more) births | 1.37 | 0.44 |
| Ind1 | Placenta previa | 0.24 | 0.63 |
| Ind2 | Ablatio | 0.46 | 0.61 |
| Ind3 | Disproportion | 2.51 | 0.36 |
| Ind4 | Pre-eclampsia | 1.76 | 0.51 |
| Ind5 | Dystocia | 6.94 | 0.77 |
| Ind6 | Fetal distress | 5.86 | 0.36 |
| Ind7[a] | Multiple births | 1.10 | 0.43 |
| Ind8 | Malpresentation | 7.10 | 0.19 |
| Ind9 | Psychosocial ind | 2.36 | 0.50 |
| Ind10 | Other indication | 0.83 | 0.96 |

Abbreviations: CV = coefficient of variation; ind = indication.
[a] *Twins* and *Ind7* both measure multiple births (twins or more). However, while *twins* measure all cases, *Ind7* measures only those cases where it is the main indication.

6, multiple births = 7, malpresentation = 8, psychosocial indication = 9, prolonged delivery = 10, other indication = 11).

Based on this database, we defined the variables shown in Table 1. Besides the traditional outcome measure, perinatal mortality, defined as "percentage of stillbirths and newborns dead within first week," we also used a morbidity measure, rate of asphyxia, defined as "percentage of newborns with Apgar score less than 7 at five minutes." The independent variable in focus is the cesarean section rate. The variables age, twins, and indications 1–10 are intended to measure case-mix. Cesarean section rate is assumed to be negatively correlated with perinatal mortality and rate of asphyxia, while all case-mix variables are assumed to be positively correlated with the outcome measures.

Table 2 contains data on average values and coefficients of variation (CV) of perinatal mortality, rate of asphyxia, and cesarean section rates in Sweden from

**Table 2.** Perinatal Mortality, Rate of Asphyxia, and Cesarean Section Rate, 1988–92; Percent, Average Values, and Coefficient of Variation

| | Perinatal mortality | | Rate of asphyxia | | Cesarean section rate | |
|---|---|---|---|---|---|---|
| Year | Average | CV | Average | CV | Average | CV |
| 1988 | 0.60 | 0.45 | 1.29 | 0.32 | 10.89 | 0.18 |
| 1989 | 0.62 | 0.44 | 1.23 | 0.35 | 10.70 | 0.17 |
| 1990 | 0.61 | 0.41 | 1.03 | 0.36 | 10.73 | 0.18 |
| 1991 | 0.64 | 0.36 | 0.95 | 0.34 | 10.82 | 0.19 |
| 1992 | 0.50 | 0.50 | 0.92 | 0.46 | 10.73 | 0.25 |
| 1988–92 | 0.59 | 0.44 | 1.08 | 0.39 | 10.74 | 0.20 |

Abbreviation: CV = coefficient of variation.

1988–92. Obviously, the variation among hospitals has been large and persisting, especially regarding the outcome measures. This is confirmed by historical data. The average cesarean section rate increased rapidly during the 1970s, while the perinatal mortality and the rate of asphyxia decreased. Superficially, this seems to indicate that the increase in cesarean sections contributed to improved medical outcome. From the early 1980s on, however, the cesarean section rate stabilized and even showed a slight decrease, while the perinatal mortality and the rate of asphyxia continued to decrease. Notwithstanding, the variation among hospitals has remained large (14).

## ESTIMATION METHODS

Obstetrical departments with higher cesarean section rates should, other things being equal, have better health outcomes in terms of lower perinatal mortality and/or lower rates of asphyxia. In other words, if we could find two obstetrical departments with exactly the same case-mix, we would expect the one with the higher cesarean section rate to have the best health outcome. Otherwise, if no such relationship can be established, the lower cesarean section rate would be optimal.

We first specify the following regression equations for health outcome, $HO_{i,t}$ (perinatal mortality or rate of asphyxia), at department i, in a specific time period, t (year):

$$HO_{i,t} = \beta_{1,t} + \beta_{2,t} * CSR_{i,t} + \beta_{3,t} * X_{3i,t} + ..... + \beta_{14,t} * X_{14i,t} + \epsilon_{i,t} \qquad (i = 1,2,.....,59)$$

$$(t = 1988–92)$$

where $\beta_{1,t}$ are the intercepts, $\beta_{2,t}$...$\beta_{14,t}$ the coefficients, $CSR_{i,t}$ the cesarean section rates, and $X_{3i,t}$,...,$X_{14i,t}$ a set of variables measuring case-mix, i.e., age, twins, and indications 1–10[5] (defined in Table 2). The stochastic disturbance terms, $\epsilon_{i,t}$, are assumed to be normally distributed with mean zero and variance $\sigma_i^2$, where $\sigma_i^2 = \sigma^2/N_i$, and $N_i$ is the number of newborns at department i. By controlling for case-mix, we could, theoretically, eliminate the effect of a confounding factor that influences both the dependent and the independent variable. We test the null hypothesis, $\beta_{2,t} = 0$, against the alternative, $\beta_{2,t} < 0$. Note that a negative sign implies a positive effect on health outcome.

We estimate the models equation by equation by use of weighted least squares (WLS), and test for poolability in order to gain in degrees of freedom if pooling is appropriate and to assess the stability of the estimated regressions. The weighting variable is the reciprocal of the variance, $w_i = 1/(N_i)^2$.

We test for poolability by use of conventional F tests (Chow test) for non-normality of the residuals by Jarque-Bera's skewness-kurtosis test, and for homoscedasticity by the Breusch-Pagan test. Functional form misspecification is tested by the Ramsey RESET test (1;9;13). To test for robustness of the results, we reestimate the models using 5-year average values of all variables. We also estimate the models excluding all but two case-mix variables. These additional tests are discussed below.

## RESULTS

The results of the regression analyses are summarized in Table 3 (perinatal mortality) and Table 4 (rate of asphyxia). Comparison between different years clearly demonstrates instability in the coefficients for cesarean section rate and for the case-mix variables. In both models, the coefficient for cesarean section changes sign as well as magnitude between periods. However, it does not differ significantly from zero except in one period (perinatal mortality, 1992). Thus, the findings do

not indicate that higher cesarean section rates result in lower perinatal mortality or lower rates of asphyxia.[6]

The Jarque-Bera test (J-B), applied to test for normality in residuals, showed that in most cases the null hypothesis could not be rejected at the 10% level. For two models,[7] however, it was rejected at the 1% level. From the Breusch-Pagan test (B-P) it appeared that most of the models showed heteroscedastic residual variance.[8] The RESET test indicates functional form misspecification for three models.[9]

We also ran WLS regressions of reduced models with two control variables only, age and twins, which—in contrast to indications 1–10—are objectively measurable. This change influences the coefficients for cesarean section, with respect to sign as well as significance. There is a significant[10] positive effect of cesarean section rate on perinatal mortality in two single-year cross-section regressions (1988, 1990) and in the pooled regression (1988–92), and on rate of asphyxia in the single-year cross-section regression for 1988. Otherwise, the main difference was that the coefficient for twins became significantly different from zero in most periods. The hypothesis that all coefficients excluded in the reduced models are equal to zero (slope restrictions) was tested using a joint F test, and was rejected for all but three periods (perinatal mortality, 1988; rate of asphyxia, 1988 and 1991). Detailed estimation results are available from the authors.

An assessment of multicollinearity is presented in Table 5.[11] The first column reports the internal coefficients of determination for the regressor variables. The remaining columns contain coefficients of correlation between the regressor variables in the pooled regression equation. The highest variance inflation factor (VIF) for the regressors included is 3.846.[12] A rule of thumb is that a VIF greater than 10 indicates strong multicollinearity (12). Our conclusion is that the individual $t$ ratios should not be too much affected by multicollinearity. However, the correlation coefficient between the variables twins and indication 7 is high, 0.82, and the internal $R^2$s for twins and indication 7 is 0.68 and 0.74, respectively. This made us re-estimate the models excluding either of these variables. However, this did not change the results very much. The coefficient for cesarean section rate changed from negative to positive in 1988 for perinatal mortality, and became significantly positive in 1992 for rate of asphyxia.

The Chow tests for equal slopes across equations show that the null hypotheses could be rejected at the 1% level for perinatal mortality, $F(52;45) = 3.23$, and at the 5% level for rate of asphyxia, $F(52;45) = 1.79$. Tests for equal slopes *and* equal intercepts across equations show that the null hypotheses could be rejected at the 1% level for both perinatal mortality, $F(56;45) = 3.53$, and rate of asphyxia, $F(56;45) = 3.61$, which indicates that pooling is not justified in this case. Moreover, pooling does not change the results.[13]

## DISCUSSION AND CONCLUDING REMARKS

In the introduction, we stated that only a minor part of the variation in cesarean section rates among obstetrical departments in Sweden is explained by differences in demand- and supply-related variables. This indicates inefficiency. Some departments perform too many cesareans and some other departments too few. To find out which rate is optimal, we have to estimate the health production function, i.e., the effect of cesarean section rate on health outcome. The objective of this study was to estimate the effect of cesarean section rate on health outcome in terms of perinatal mortality and rate of asphyxia, respectively.

**Table 3.** Parameter Estimates of Perinatal Mortality: Equation-by-Equation WLS and Pooled WLS; Coefficients (*t* values) for Regressors

| Variable | 1988 | 1989 | 1990 | 1991 | 1992 | Pooled |
|---|---|---|---|---|---|---|
| Constant | 0.777E-01 | 0.919E-01 | −0.665[b] | 0.546[c] | 0.360 | 0.147 |
| | (0.201) | (0.221) | (−2.378) | (1.469) | (1.135) | (1.080) |
| CSR | −0.409E-02 | 0.205E-01 | 0.847E-02 | 0.848E-02 | −0.225E-01[c] | −0.224E-02 |
| | (−0.165) | (0.722) | (0.453) | (0.584) | (−1.418) | (−0.238) |
| Age | 0.700E-01[b] | 0.188E-01 | 0.383E-01[b] | 0.87E-02 | 0.126E-01 | 0.385E-01[a] |
| | (2.333) | (0.651) | (1.974) | (0.496) | (0.601) | (3.573) |
| Twins | −0.201 | 0.957E-01 | −0.344[a] | 0.101 | 0.713E-01 | −0.113[a] |
| | (−0.846) | (0.876) | (−3.854) | (0.727) | (1.064) | (−2.950) |
| Ind1 | 0.277E-01 | −0.335 | 0.166 | 0.288[b] | 0.235 | −0.178E-01 |
| | (0.075) | (−0.748) | (0.953) | (1.960) | (1.219) | (−0.170) |
| Ind2 | −0.272E-01 | 0.303[c] | 0.29[b] | −0.217E-01 | 0.137 | 0.136[b] |
| | (−0.145) | (1.503) | (2.092) | (−0.231) | (1.166) | (2.193) |
| Ind3 | 0.479E-01 | 0.507E-02 | 0.492E-01[c] | −0.231E-01 | −0.858E-01[a] | −0.299E-02 |
| | (1.081) | (0.088) | (1.604) | (−0.686) | (−3.010) | (0.168) |
| Ind4 | 0.449E-02 | 0.163[a] | −0.669E-01[b] | −0.659E-01[b] | −0.457E-01 | 0.178E-01 |
| | (0.098) | (3.007) | (−2.204) | (−1.917) | (−1.089) | (−0.912) |
| Ind5 | −0.912E-02 | −0.861E-02 | −0.351E-02 | 0.182E-01[a] | 0.105E-02 | −0.601E-02[c] |
| | (−0.758) | (−0.692) | (−0.507) | (2.777) | (0.176) | (−1.391) |
| Ind6 | −0.583E-01[b] | −0.358E-01[c] | 0.227E-01[c] | 0.410E-01[a] | −0.294E-01[c] | −0.209E-01[b] |
| | (−2.263) | (−1.621) | (1.431) | (2.518) | (−1.624) | (−2.252) |
| Ind7 | 0.703[a] | 0.137 | 0.301[b] | 0.216 | 0.146[c] | 0.338a |
| | (2.536) | (0.830) | (2.388) | (1.209) | (1.489) | (6.336) |
| Ind8 | −0.546E-01 | −0.227E-01 | 0.620E-01[a] | −0.109[a] | 0.412E-01 | −0.837E-02 |
| | (−1.378) | (−0.443) | (2.481) | (−2.479) | (1.094) | (−0.670) |
| Ind9 | −0.692E-01 | −0.185E-02 | 0.671E-01[b] | 0.509E-03 | −0.315E-01 | 0.408E-02 |
| | (−0.901) | (−0.027) | (2.347) | (0.016) | (1.111) | (0.200) |
| Ind10 | −0.743E-02 | −0.581E-01 | 0.150E-01 | 0.124a | 0.158E-01 | 0.406E-02 |
| | (−0.111) | (−1.044) | (0.380) | (3.274) | (−0.506) | (0.200) |

(*Continued*)

**Table 3.** *(Continued)*

| Variable | 1988 | 1989 | 1990 | 1991 | 1992 | Pooled |
|---|---|---|---|---|---|---|
| Evaluation: | | | | | | |
| *df* | 45 | 45 | 45 | 45 | 45 | 281 |
| *R* | 0.615 | 0.476 | 0.753 | 0.614 | 0.562 | 0.281 |
| Standard error | 0.25 | 0.29 | 0.17 | 0.18 | 0.19 | 0.28 |
| J-B $\chi^2$ (2) | 0.534 | 0.230 | 0.902 | 11.438[a] | 0.308 | 3.301 |
| B-P $\chi^2$ (13) | 14.07 | 34.47[a] | 23.77[b] | 39.70[a] | 24.12[b] | 49.16[a] |
| RESET F(3;42) | 4.640[a] | 0.726 | 0.914 | 2.404 | 1.053 | 9.256[a] |

Abbreviations: WLS = weighted least squares; CSR = cesarean section rate; J-B = Jarque-Bera test; B-P = Breusch-Pagan test; ind = indication.
[a] Significant at the 1% level.
[b] Significant at the 5% level.
[c] Significant at the 10% level.
All tests are one-tailed.

**Table 4.** Parameter Estimates of Rate of Asphyxia: Equation-by-Equation WLS and Pooled WLS; Coefficients (*t* values) for Regressors

| Variable | 1988 | 1989 | 1990 | 1991 | 1992 | Pooled |
|---|---|---|---|---|---|---|
| Constant | 0.872[c] | 1.833[a] | −0.659 | 0.846 | 1.039[b] | 0.956[a] |
| | (1.581) | (2.866) | (−1.133) | (1.288) | (2.105) | (4.355) |
| CSR | 0.307E-01 | −0.191E-01 | 0.124E-01 | 0.497E-02 | 0.177E-01 | 0.851E-02 |
| | (0.868) | (−0.438) | (0.318) | (0.194) | (0.714) | (0.558) |
| Age | 0.242E-01 | −0.101[b] | 0.119E-01 | −0.161E-01 | −0.724E-01[b] | −0.379E-02 |
| | (0.567) | (−2.274) | (0.296) | (−0.517) | (−2.226) | (−0.218) |
| Twins | −0.179 | 0.514E-01 | 0.159E-01 | 0.116 | 0.225[b] | −0.100[c] |
| | (−0.531) | (0.305) | (0.085) | (0.472) | (2.162) | (−1.611) |
| Ind1 | −0.557E-01 | 0.152 | 0.485[c] | −0.929E-01 | 0.938[a] | 0.261[c] |
| | (−0.106) | (0.220) | (1.336) | (−0.357) | (3.131) | (1.544) |
| Ind2 | −0.270E-01 | −0.861E-01 | −0.463E-01 | −0.179E-01 | 0.133 | −0.124 |
| | (−0.102) | (−0.277) | (−0.160) | (−0.108) | (0.724) | (−1.237) |
| Ind3 | −0.348E-01 | −0.129[c] | −0.175E-02 | −0.139[b] | −0.164[a] | −0.236E-01 |
| | (−0.551) | (−1.452) | (−0.027) | (−2.339) | (−3.694) | (−0.824) |
| Ind4 | 0.727E-01 | −0.220[a] | −0.141E-01 | −0.179E-01 | −0.164[a] | −0.325E-01 |
| | (1.119) | (−2.620) | (−0.224) | (−0.295) | (−2.508) | (−1.035) |
| Ind5 | 0.262E-01[c] | 0.340E-01[b] | −0.226E-02 | 0.113E-01 | 0.222E-01[b] | 0.915E-02[c] |
| | (1.526) | (1.774) | (−0.157) | (0.970) | (2.380) | (1.312) |
| Ind6 | −0.693E-01[b] | 0.286E-01 | 0.634E-01[b] | 0.248E-01 | −0.160E-02 | 0.158E-01 |
| | (−1.889) | (0.842) | (1.923) | (0.863) | (−0.057) | (1.058) |
| Ind7 | 0.931[b] | 0.319 | −0.280 | 0.258 | −0.204E-01 | 0.478[a] |
| | (2.358) | (1.250) | (−1.067) | (0.819) | (−0.133) | (5.542) |
| Ind8 | −0.714E-01 | 0.163[b] | 0.105[b] | 0.143E-02 | 0.351E-01 | −0.473E-01[a] |
| | (−1.265) | (2.069) | (2.011) | (0.018) | (0.599) | (−2.347) |
| Ind9 | −0.668E-01 | −0.129 | 0.220[a] | −0.249E-01 | 0.603E-01[c] | 0.185E-01 |
| | (−0.610) | (−1.237) | (3.707) | (−0.456) | (1.368) | (0.560) |
| Ind10 | −0.763E-01 | −0.613E-01 | −0.321E-01 | −0.451E-01 | −0.795E-01[c] | −0.462E-01[c] |
| | (−0.798) | (−0.714) | (−0.390) | (−0.674) | (−1.631) | (−1.412) |

(*Continued*)

**Table 4.** *(Continued)*

| Variable | 1988 | 1989 | 1990 | 1991 | 1992 | Pooled |
|---|---|---|---|---|---|---|
| Evaluation: | | | | | | |
| *df* | 45 | 45 | 45 | 45 | 45 | 281 |
| *R* | 0.559 | 0.458 | 0.496 | 0.451 | 0.675 | 0.231 |
| Standard error | 0.35 | 0.45 | 0.36 | 0.32 | 0.29 | 0.44 |
| J-B $\chi^2$ (2) | 1.088 | 0.356 | 0.929 | 1.420 | 0.177 | 11.866[a] |
| B-P $\chi^2$ (13) | 38.18[a] | 22.18[c] | 14.92 | 15.38 | 30.63[a] | 43.16[a] |
| RESET F(3;42) | 3.890[c] | 0.115 | 0.957 | 2.236 | 0.890 | 0.057[a] |

For abbreviations, see Table 3.
[a] Significant at the 1% level.
[b] Significant at the 5% level.
[c] Significant at the 10% level.
All tests are one-tailed.

**Table 5.** Assessment of Multicollinearity by Internal Coefficients of Determination and the Correlation Matrix of Regressors

| | | $R^2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CSR | 0.36 | 1 | | | | | | | | | | | | |
| 2 | Age | 0.30 | .35 | 1 | | | | | | | | | | | |
| 3 | Twins | 0.68 | .22 | .30 | 1 | | | | | | | | | | |
| 4 | Ind1 | 0.24 | .32 | .27 | .30 | 1 | | | | | | | | | |
| 5 | Ind2 | 0.26 | .17 | .15 | .19 | .27 | 1 | | | | | | | | |
| 6 | Ind3 | 0.24 | .33 | *.01* | −.19 | *−.01* | *.05* | 1 | | | | | | | |
| 7 | Ind4 | 0.20 | *.09* | *.10* | .30 | .23 | .30 | *−.02* | 1 | | | | | | |
| 8 | Ind5 | 0.14 | *.00* | .22 | .12 | *.03* | *−.02* | −.12 | .16 | 1 | | | | | |
| 9 | Ind6 | 0.23 | *.01* | .25 | .22 | .15 | .34 | *.06* | .18 | *.05* | 1 | | | | |
| 10 | Ind7 | 0.74 | .28 | .39 | .82 | .38 | *.27* | −.17 | .30 | .14 | .25 | 1 | | | |
| 11 | Ind8 | 0.31 | .19 | .16 | .43 | .21 | .33 | *−.04* | .29 | *.04* | .26 | .49 | 1 | | |
| 12 | Ind9 | 0.18 | .14 | .24 | .17 | *.06* | *.02* | −.16 | *.09* | .28 | *.02* | .28 | *.07* | 1 | |
| 13 | Ind10 | 0.05 | *.05* | *−.10* | *−.03* | *.05* | *−.03* | *−.03* | *−.03* | *.01* | −.14 | *−.07* | *.01* | *−.02* | 1 |

Abbreviations: CSR = cesarean section rate; ind = indication.
Figures in italics indicate nonsignificance at the 5% level.

An obstetrical department may have a high cesarean section rate because of a high prevalence of complicated deliveries, which makes it necessary or desirable to perform relatively many cesarean sections. When controlling for case-mix, however, we should expect to find a negative correlation between outcome (perinatal mortality, rate of asphyxia) and cesarean section rate. As shown above, this is generally not the case. The null hypothesis of a zero effect could not be rejected: the findings did not indicate that a higher cesarean section rate implies better health outcome in terms of perinatal mortality or rate of asphyxia. Thus, our results are in accordance with those of Joffe et al. (10), who found the same lack of correlation using a different methodological approach.

A critical issue in our analysis is whether the outcome and case-mix measures are appropriate. Regarding the outcome measures, we chose perinatal mortality and rate of asphyxia, both of which seem to be valid and readily available from the Medical Birth Registry (14;15). To be sure, these measures are relatively crude. Still, they are the only available measures suitable for comparisons across departments. Ideally, we should use a broader range of outcome measures, preferably combined into a health outcome index. With regard to case-mix, we tried alternative measures (reduced models), but these did not change the results.

The instability of the models is an obvious problem. The coefficients for cesarean section rate as well as for the case-mix variables change sign and magnitude from year to year. The reason for this is not clear. One possible explanation could be the fact that we deal with low-incidence events. However, the fact that several coefficients are significantly different from zero indicates that the observed variation is larger than would be expected by chance alone. This was confirmed by chi-square tests, showing that the hypothesis of no variation in the dependent variables could be rejected.

Another conceivable explanation has to do with the statistical power of the test. The fact that the sample is relatively small implies a high probability of type II error, i.e., not rejecting the null hypothesis when it is in fact false. By expanding the sample size, we could in principle reduce the variance of an estimator. However, it is not quite clear whether this would influence the problem with changing signs of the regressor coefficients.

Alternatively, the instability may indicate that the models are misspecified due to omission of variables or nonlinearity. The RESET test indicates functional form misspecification for some of the models. To check for nonlinearity, we standardized the dependent variables in the reduced models with respect to the variables age and twins, and looked at scatter diagrams showing the correlation between cesarean section rate and the standardized outcome variables. These inspections did not indicate the existence of either nonlinearity or linearity.

Assuming that our conclusion about inefficiency due to variation in cesarean section rates is correct, a natural next step is to estimate the economic consequences of this inefficiency, i.e., to measure the welfare loss from undesired medical practice variation (4;8). Since we did not find any significant effect of cesarean section rate on health outcome, there is in principle no reason for performing cesarean section rates above the minimum level (adjusted for case-mix). However, with regard to the problems discussed above, there is good reason for caution in assessing the validity of our results.

If we assume that, in fact, there is a significant positive effect of cesarean section rate on health outcome, we are back to the problem of estimating the optimal rate as a basis for estimating the welfare loss from undesired variation. In our previous

study, we tried to do so, using a model developed by Phelps and Parente (17). One of the problems that remains to be solved is to estimate the marginal value (demand) curve for cesarean sections. Since there is no real market for health care in Sweden, there is no self-evident method to estimate the demand. Possibly, we could use the contingent valuation approach, e.g., by asking patients who are about to go through a delivery about their willingness to pay for (valuation of) a cesarean section, given specified risks and benefits (11).

Finally, we should also look closer at available policy options, i.e., the opportunities for changing practice towards the best practice as defined by science and proven experience. It is a well-known fact that it is difficult to influence practice by information alone (6). However, we are convinced that improved knowledge and information about the economic consequences of different medical practices would be a valuable basis for influencing practice.

## NOTES

[1] More than two-thirds of cesarean sections are performed exclusively for the newborns' sake (15).

[2] The outcome measures used were: perinatal mortality, Apgar scores at 1 and 5 minutes, onset of respiration after 1 minute, postnatal transfusion, postnatal infection, thromboembolism, low hemoglobin concentration at discharge, and puerperal psychosis.

[3] This is contradicted by Rydhström et al. (18), who found no such correlation in twin deliveries.

[4] Apgar score is a method of judging the condition of a newborn baby. The baby is given a maximum of two points on each of five criteria: color of the skin, heartbeat, breathing, muscle tone and reaction to stimuli.

[5] Indication 11, "other indication," is a residual, and is not included in the analysis.

[6] Since perinatal mortality as well as rates of asphyxia are generally very low in Sweden, they are extremely sensitive to random variation. Therefore, we also ran WLS regressions based on average values of all variables in the 5-year period. This did not change the results concerning the coefficient for cesarean section rate, but gave a higher $R^2$ (0.792 for perinatal mortality and 0.670 for rate of asphyxia).

[7] 1991 for perinatal mortality and the pooled model for rate of asphyxia.

[8] We also estimated $t$ values with the covariance estimator, which allows for heteroscedasticity, but this did not change the results (13).

[9] 1988 and pooled model for perinatal mortality, and 1988 for rate of asphyxia.

[10] At the 5% level for perinatal mortality in 1990, otherwise at the 10% level.

[11] Table 5 is based on the pooled data, but the results are similar for every year.

[12] The VIF for a regressor $X_k$ is calculated as $(1-R_k^2)^{-1}$, where $R_k^2$ is the $R^2$ from regressing the $k$th independent variable on all the other independent variables.

[13] We also estimated the five equations as a set of linear equations, using Zellner's seemingly unrelated regressions (SUR, or error-related regressions), i.e., one set of equations with perinatal mortality as dependent variable and the other with rate of asphyxia as dependent variable. The disturbances across the regressions were allowed to be freely correlated. Equation-by-equation OLS were used to obtain an estimate of the disturbance covariance matrix $\Omega$ and then generalized least squares (GLS) was used on the "stacked" set of equations (19). However, neither did the SUR estimations change the main conclusion of the WLS regressions.

## REFERENCES

1. Breusch, T., & Pagan, A. A simple test for heteroscedasticity and random coefficient variation. *Econometrica,* 1979, 47, 1287–94.
2. Broadhead, T. J., & James, D. K. Worldwide utilization of cesarean section. In B. L. Flamm & E. J. Quilligan (eds.), *Cesarean section: Guidelines for appropriate utilization.* New York: Springer-Verlag, 1995.

3. Chalmers, I., Enkin, M., & Keirse, M. *Effective care in pregnancy and childbirth. vol. 2.* Oxford: Oxford University Press, 1991.
4. Clark, L., Mugford, M., & Paterson, C. How does the mode of delivery affect the cost of maternity care? *British Journal of Obstetrics and Gynaecology,* 1991, 98, 519–23.
5. Eckerlund, I., & Gerdtham, U.-G. *Econometric analysis of variation in cesarean section rates: A cross-sectional study of 59 obstetrical departments in Sweden. International Journal of Technology Assessment in Health Care*, 1998, 14, 774–87.
6. Evans, R. G. The dog in the night: Medical practice variations and health policy. In T. F. Andersen & G. Mooney (eds.), *The challenges of medical practice variations.* London: Macmillan, 1990.
7. Flamm, B. L., & Quilligan, E. J. (eds). *Cesarean section: Guidelines for appropriate utilization.* New York: Springer-Verlag, 1995.
8. Gardner, L. B. Economic considerations in cesarean section use. In B. L. Flamm & E. J. Quilligan (eds.), *Cesarean section: Guidelines for appropriate utilization.* New York: Springer-Verlag, 1995.
9. Jarque, C. M., & Bera, A. K. A Test for normality of observations and regression residuals. *International Statistical Review,* 1987, 55, 163–72.
10. Joffe, M., Chapple, J., Paterson, C., & Beard, R. W. What is the optimal cesarean section rate? An outcome based study of existing variation. *Journal of Epidemiology and Community Health,* 1994, 48, 406–11.
11. Johansson, P.-O. *Evaluating health risks: An economic approach.* Cambridge: Cambridge University Press, 1995.
12. Kennedy, P. *A guide to econometrics,* 3rd ed. Oxford: Blackwell, 1992.
13. Kmenta, J. *Elements of econometrics.* 2nd ed. New York: Macmillan, 1990.
14. *Medical birth registration.* Stockholm: National Board of Health and Welfare, 1988–92.
15. Nielsen, T. F., Otterblad Olausson, P., & Ingemarsson, I. The cesarean section rate in Sweden: The end of the rise. *Birth,* 1994, 21, 1.
16. O'Driscoll, K., & Foley, M. Correlation of decrease in perinatal mortality and increase in CS rates. *Obstetrics and Gynecology*, 1983, 61, 1–5.
17. Phelps, C. E., & Parente, S. T. Priority setting for medical technology and medical practice assessment. *Medical Care,* 1990, 28, 703–23.
18. Rydhström, H., Ingemarsson, I., & Ohrlander, S. Lack of correlation between a high CS rate and improved prognosis for low-birthweight twins (<2500 g). *British Journal of Obstretrics and Gynaecology*, 1992, 97, 229–33.
19. Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association,* 1962, 57, 348–68.

# Benchmarking in obstetric care – a comparative study based on data envelopment analysis (DEA)

Ingemar Eckerlund & Magnus Tambour

**Abstract**

Benchmarking is increasingly used in performance monitoring of the health care sector, for self-assessment as well as external accountability – in order to promote efficiency. However, a prerequisite for achieving this objective is that the benchmarking process is based on scientifically sound methods. Using DRG-data from 19 Swedish hospitals in 1994-1995, we apply data envelopment analysis (DEA) to compare technical efficiency (productivity) of inpatient obstetric care. The results indicate an average potential for enhanced productivity of at least six-seven percent each year. Most hospitals show decreasing productivity between 1994 and 1995, partly due to their inability to adjust production capacity and costs to decreasing demand. From a policy perspective, DEA could be very useful in performance monitoring. An important advantage of the DEA method is that it helps identify the strengths and weaknesses of the various units (hospital departments) studied. We conclude that there are good reasons for adopting the DEA method for benchmarking in other types of hospital departments as well, and preferably on a larger scale, i.e., with more observations, and also for including health outcome and quality indicators among the output variables.

# 1. INTRODUCTION

Benchmarking, i.e., a comparison of the performance of similar production units, has become common in the health care sector. The idea is to stimulate a process of learning from "the best" and thereby promote efficiency. For several reasons, it is difficult to assess the overall efficiency of health care provision, i.e., the relation between health outcomes and costs at an organisational level, such as hospital or hospital department. The most important difficulty seems to be the lack of valid, reliable and generally accepted measures of the outcome or effects of health care. These must be assessed in relation to the ultimate goal of the activity studied, i.e improvements in life expectancy, and quality of life for the patient. Furthermore, isolating the effect of a specific medical treatment or surgical procedure may be difficult, since it often takes some time before any effect can be observed. This makes it difficult to separate the effect of the intervention from the natural history of the disease and the effects of other activities in the period. The same kind of validity problems is to be found on the resource side, since it may be difficult to define, separate and measure the relevant costs.

Generally, *technical efficiency* or *productivity*[1], defined as the ratio between output and input of resources, is easier to assess. However, productivity measurement is not an easy task when – as in health care provision – multiple outputs are produced by multiple inputs. Substituting productivity assessment for overall efficiency assessment presupposes an unambiguous and positive correlation between the output measure used and the goal of the activity.

Frequent output measures in health care are *outpatient visits, admissions, discharges, examinations (diagnostic tests)* and *surgical procedures*. These measures may be used to describe the volume of a certain activity, but cannot be used to analyse productivity or efficiency without further considerations. For example, fewer outpatient visits per physician hour (lower technical efficiency) may well result in improved effectiveness, and even lower costs. If the doctor spends more time per patient/visit, the treatment goal may be reached with a smaller total number of visits. This example demonstrates a potential conflict between productivity (measured in this way) and overall efficiency. However, this is not necessarily the case, provided that adequate productivity measures are used. In principle, comparisons of productivity over time or between various production units may be valid if the content and quality of the outputs are constant between the times of measurement or if the outputs from the units compared are equivalent in a cross-section study. If not, we must adjust for the differences. Many studies of health care productivity do not fulfil these prerequisites. Moreover, traditional output measures only reflect limited aspects of the activities and/or do not correctly link outputs and inputs (costs).

In the light of these problems, an alternative method for productivity measurement, Data Envelopment Analysis (DEA), has gained interest within health economics (1, 2, 3, 4, 5). DEA is a non-parametric method based on linear programming techniques. Using observed data on production volumes and resources/costs, DEA can be applied to derive a "best practice frontier", consisting of optimal combinations of inputs and outputs. The individual producer's technical efficiency level is determined by the distance to the estimated best practice frontier. An important advantage of DEA is that several inputs and outputs can easily be included, provided that enough observations (e.g., hospital departments) are included in the dataset. Furthermore, the variables do not have to be measured in the same units of measurement or weighted (between themselves) beforehand. Hence, it is possible, albeit not optimal, to si-

[1] In the data envelopment analysis (DEA) literature, *productivity* often denotes changes in technical efficiency over time. However, here *productivity* and *technical efficiency* are used as synonyms.

multaneously use both quantities and costs as measures of input, i.e., the DEA estimation technique is independent of the unit of measurement.

The purpose of this study is to benchmark obstetric care in Swedish hospitals by comparing their performance in terms of technical efficiency (productivity). We limit this study to include only *inpatient obstetric care*, i.e., inpatient maternity care and deliveries. This choice is based upon the judgement that resources and costs as well as output can be reasonably well and completely described for that sector of care, even if some resources are shared with outpatient care. Furthermore, there is a strong tradition of performance monitoring in obstetric care in Sweden, concerning the cesarean section rate, perinatal mortality, etc, and a great interest in new methods for monitoring (6).

Section 2 presents the data and the variables selected. The method applied, Data Envelopment Analysis (DEA), is described in section 3, and the results are presented in section 4. The paper ends, in section 5, with a discussion and some concluding remarks.

## 2. MATERIAL

**Data**
The data were collected by a questionnaire sent to all 58 hospitals with separate departments of obstetrics and gynecology in Sweden. The information asked for was outputs in terms of discharges according to Diagnosis Related Groups (DRG) and resource use and costs of inpatient obstetrical care in 1994 and 1995. The reason for using DRG as the output measure is that it is the best measure available, even if it has not yet been adopted throughout all hospitals.

25 hospitals answered the questionnaire, while 16 reported that they did not use the DRG system. The remaining 17 hospitals did not answer despite reminders. However, most of these did not use DRG. Six of the 25 answering hospitals were excluded due to incomplete answers. Thus, 19 hospitals (33 percent of all hospitals surveyed, but close to 100 percent of those using DRG) are included in the study; five university hospitals, ten well-equipped general hospitals and four other (local) hospitals. This means some overrepresentation of large hospitals, which is evident from the fact that the study includes about 40 percent of all deliveries in the country (6).

**Variables**
According to a common rule of thumb in DEA, the number of observations should be at least three times the number of variables, i.e., the sum of inputs and outputs (4). Thus, the number of variables in this study should not exceed six ($19 > 3x6$). The various types of outputs (DRGs) asked for in the questionnaire were grouped into the following three categories:

- *cesarean sections (DRG 370, 371)*
- *vaginal deliveries (DRG 372, 373)*
- *other DRGs.*

Two input variables were derived from the questionnaire, i.e. physician costs and costs for other staff. Since there were only very small wage differences among hospitals within the various personnel categories at this time, these variables probably reflect resource input fairly

well. A variable reflecting the supply of beds in maternity wards was also included (6). Thus, the following input variables were included in the analysis:

- *physician costs*
- *costs for other staff*
- *number of beds.*

The variables are presented in Table 1.

**Table 1. Variables in the DEA-study; min-, median and max-values (1994 and 1995).**

| Variable | Year | Min | Median | Max |
|---|---|---|---|---|
| ***Outputs:*** | | | | |
| Cesarean sections | *1994* | 68 | 206 | 761 |
| | *1995* | 52 | 204 | 558 |
| Vaginal deliveries | *1994* | 634 | 1 838 | 4 216 |
| | *1995* | 536 | 1 706 | 3 793 |
| Other DRGs | *1994* | 92 | 313 | 607 |
| | *1995* | 85 | 244 | 463 |
| ***Inputs:*** | | | | |
| Physician costs (1000 SEK) | *1994* | 420 | 2 967 | 8 655 |
| | *1995* | 498 | 2 999 | 8 545 |
| Cost for other staff (1000 SEK) | *1994* | 5 056 | 18 808 | 45 000 |
| | *1995* | 4 701 | 19 098 | 43 000 |
| Number of beds | *1994* | 12 | 35 | 70 |
| | *1995* | 10 | 38 | 59 |

Obviously, vaginal deliveries are the most common type of output. Among single diagnosis groups *DRG 373 Vaginal delivery without complicating diagnoses* is the most frequent by far, covering 60-70 percent of all discharges from these departments.

**Data quality**

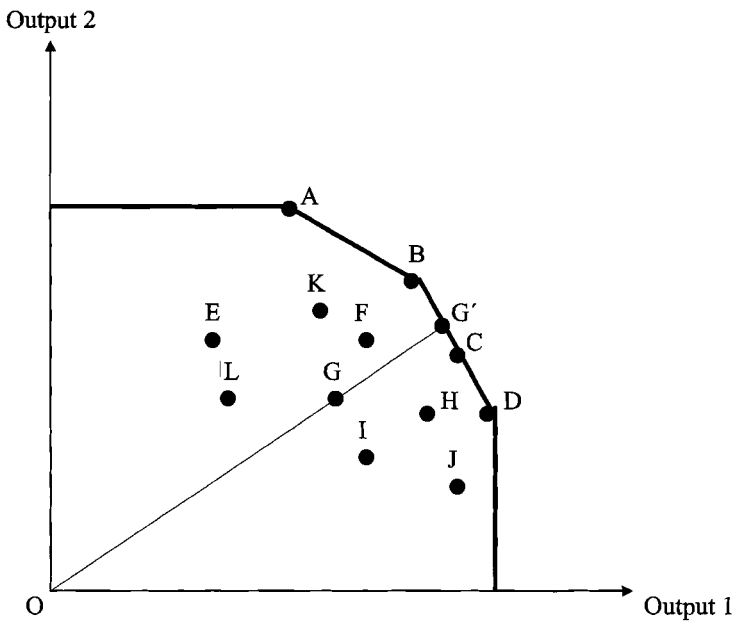
Output data are of high quality regarding the number of *cesarean sections* and *vaginal deliveries*. The output group *other DRGs* is probably less homogenous, mainly due to a potential variation in registration practice. As a rule, physicians share their time between various activities. Thus, data on *physician costs* are based on estimations of their average time distribution. The variable *costs for other staff* is more straightforward, even if some hospitals reported difficulties in separating the costs referring to inpatient care.

# 3. METHODS

**Efficiency analysis**

Data Envelopment Analysis (DEA) uses mathematical optimisation techniques (linear programming) to calculate the relative efficiency level of various production units (1, 10). In principle, this is executed by comparing every single unit with the most (technically) efficient units. Figure 1 demonstrates the principles behind DEA by a simplified example with only two types of outputs.

**Figure 1. Production possibility frontier; DEA**




Assume a number of production units (A - L) which all produce two types of output (output 1 and 2). For illustration purposes only, we also assume all units to have the same amount of inputs, which are used in the production of various combinations of outputs. Units A, B, C and D are most efficient. The connection of these points constitutes a *best practice frontier*, the production possibility curve, which is equal to the set of optimum (or maximum) combinations of the two types of outputs. These units are regarded as technically efficient, since it is impossible for any of them to produce more of one output without producing less of the other. The other units are technically inefficient, since it would be possible for them to increase one or both outputs. It should be noted that in the application of DEA, the frontier is derived from the units studied only, which means that producing output combinations outside the estimated frontier may in fact be possible. In other words, the "true" frontier may be situated further from origo.

According to the DEA-method, the efficiency (productivity) of each production unit is calculated by relating the actual production level to the maximum possible production level. In

principle, unit G could produce G′. The relative productivity is denoted by the ratio OG/OG′ and the distance to the frontier is a measure of the potential production increase, given the actual production mix (the relation between output 1 and 2).

More formally, the DEA-method may be described as follows (10):

Let $x_{kn}$ denote the amount of the n′th input used by unit $k$, and let $y_{km}$ denote the amount of the $m$′th output produced by unit $k$. The technical efficiency of unit $k$ is estimated by solving the following linear programming problem[2]:

$$(1) \qquad F_i(y,x) = \min_{\lambda,z} \lambda$$

subject to:

$$(1a) \quad y_{km} \le \sum_k z_k y_{km} \quad m = 1,...M,$$

$$(1b) \quad \sum_k z_k x_{kn} \le \lambda x_{kn} \quad n = 1,...N,$$

$$(1c) \quad z_k \ge 0 \quad k = 1,....K$$

The optimization problem is to find the smallest number (factor), λ, by which each input $x$ (used by unit $k$) can be reduced, and still produce the same output. Hence, if λ is equal to one, it is not possible to reduce the input levels (and costs) and the unit is then considered technically efficient. The $z$ variables are referred to as the intensity variables. They tell us how an inefficient unit might increase its efficiency by adopting a linear combination of the production technologies of two technically efficient departments or more, the "peers". If, for example, $z_1$ and $z_2$ are greater than zero (while all other $z$ values are equal to zero), units 1 and 2 constitute the reference technology for unit $k$. Thus, in DEA, each production unit is compared with its "peers". For example, in figure 1, unit G is compared with units B and C, since G′ can be attained by a linear combination of the outputs of these units.

In the optimisation problem specified in (1), the technical efficiency is estimated for a given period (year). It is also possible to apply the DEA method on panel data to calculate productivity changes between different periods, e.g. by estimating Malmquist indices (10) which makes it possible to calculate the change in productivity by relating inputs and outputs for each production unit to its peers in different time periods. This method also enables a decomposition of "**the total productivity change**" into **frontier** (or **"technical") change**, e.g. due to a shift in the production technology, and **efficiency change,** e.g. if a production unit has moved closer to the best practice frontier between two periods.

The default DEA-model in the software programme[3] used is the so-called input-saving measure of technical efficiency, with constant returns to scale, denoted by $F_i(y,x|C)$. However, it is also possible to calculate efficiency under nonincreasing returns to scale, $F_i(y,x|N)$, and variable returns to scale, $F_i(y,x|V)$.

[2] This is the input-saving DEA model, which is most appropiate in this case. However, Figure 1 illustrates the output-oriented DEA model which, in our opinion, makes it easier to demonstrate the idea behind DEA.
[3] OnFront™ (Economic Measurement and Quality in Lund AB).

Thus, if a production unit is found to be inefficient according to the default model, it means that it deviates from constant returns to scale. This deviation is the intuition behind the measure input scale efficiency ($S_i$), defined as $Si(y,x) = F_i(y,x|C)/F_i(y,x|V)$. By computing the three efficiency measures, we can determine if a unit is scale efficient. If it is not, we can determine whether this is due to its operating at increasing returns to scale, which is the case if $F_i(y,x|N) = F_i(y,x|C)$, or decreasing returns to scale, if $F_i y,x|N) > F_i(y,x|C)$.

To test for the impact of production volume (or capacity) on efficiency (productivity), we apply ordinary least squares (OLS) regression analysis. We test the null hypothesis of a zero effect of production volume (capacity) on the efficiency level. The dependent variable is measured in terms of *super-efficiency*, a measure calculated by comparing each department with all departments but one's own, in order to separate the technically efficient units (8). Volume (capacity) is measured in three alternative ways, in terms of *number of beds*, *total number of admissions* and *number of deliveries*, respectively. To obtain more observations, we also pool the data for 1994 and 1995 in the regression analysis.

As mentioned, the DEA-method has some important advantages compared to other methods for productivity measurement. First, it is based on few assumptions about the properties of the production function. Second, it is possible to include several inputs and outputs in the model which must not necessarily be expressed in the same measurement units or weighted between themselves beforehand.

There are also some disadvantages or limitations, however. The method presupposes that extreme cases are judged to be technically efficient, and is thus sensitive to data errors. Furthermore, the results are sensitive to the number of variables included in the analysis; the number of units regarded as efficient increases with the number of variables. Thus, the above mentioned rule of thumb, that the number of observations should be at least three times the number of variables, is intended to prevent overestimation of the efficiency levels (2).

Finally, it is important to keep in mind that the DEA-method only measures relative efficiency. Even if a unit is regarded as technically efficient, there may be other units not included in the comparison which are still better, i.e., the DEA measure is biased relative to the "true" frontier.

# 4. RESULTS

**Efficiency (productivity) of inpatient obstetric care (1994 and 1995)**
Table 2 shows the results of the DEA.

**Table 2. Efficiency (productivity) in 1994 and 1995, and changes between 1994 and 1995.**

| Hospital | Eff 1994 | Farrell 1994 | Rank 1994 | Eff 1995 | Farrell 1995 | Rank 1995 | Eff change 1994-1995 | Techn change 1994-1995 | Malmquist index |
|---|---|---|---|---|---|---|---|---|---|
| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) | (j) |
| **1** | 1.00 | 1.00 | 7 | 1.00 | 1.00 | 6 | 1.00 | 0.92 | 0.92 |
| **2** | 1.00 | 1.00 | 8 | 1.00 | 1.00 | 5 | 1.00 | 0.96 | 0.96 |
| **3** | 0.94 | 1.06 | 11 | 1.00 | 1.00 | 9 | 1.06 | 0.99 | 1.05 |
| **4** | 0.87 | 1.15 | 15 | 0.98 | 1.02 | 11 | 1.13 | 0.87 | 0.98 |
| **5** | 1.00 | 1.00 | 3 | 1.00 | 1.00 | 3 | 1.00 | 0.98 | 0.98 |
| **6** | 1.00 | 1.00 | 1 | 1.00 | 1.00 | 1 | 1.00 | 1.01 | 1.01 |
| **7** | 0.94 | 1.06 | 12 | 0.98 | 1.03 | 12 | 1.04 | 0.87 | 0.91 |
| **8** | 1.00 | 1.00 | 6 | 0.99 | 1.01 | 10 | 0.99 | 0.89 | 0.88 |
| **9** | 0.82 | 1.22 | 18 | 0.89 | 1.12 | 15 | 1.09 | 0.85 | 0.93 |
| **10** | 0.76 | 1.31 | 19 | 0.73 | 1.38 | 19 | 0.96 | 0.93 | 0.89 |
| **11** | 0.96 | 1.05 | 10 | 0.87 | 1.15 | 18 | 0.91 | 1.00 | 0.91 |
| **12** | 1.00 | 1.00 | 4 | 1.00 | 1.00 | 7 | 1.00 | 0.84 | 0.84 |
| **13** | 0.88 | 1.14 | 14 | 0.88 | 1.14 | 17 | 1.00 | 0.91 | 0.91 |
| **14** | 0.92 | 1.09 | 13 | 0.91 | 1.10 | 14 | 0.99 | 0.94 | 0.92 |
| **15** | 1.00 | 1.00 | 9 | 1.00 | 1.00 | 4 | 1.00 | 1.12 | 1.12 |
| **16** | 0.87 | 1.15 | 16 | 0.94 | 1.07 | 13 | 1.08 | 0.93 | 1.00 |
| **17** | 0.87 | 1.16 | 17 | 0.89 | 1.12 | 16 | 1.02 | 0.99 | 1.02 |
| **18** | 1.00 | 1.00 | 2 | 1.00 | 1.00 | 2 | 1.00 | 0.90 | 0.90 |
| **19** | 1.00 | 1.00 | 5 | 1.00 | 1.00 | 8 | 1.00 | 0.84 | 0.84 |
| **Average** | **0.94** | **1.07** | | **0.95** | **1.06** | | **1.01** | **0.93** | **0.94** |

**Efficiency (productivity) in 1994**
Column (b) shows that in 1994, nine of the 19 hospital departments obtained a score equal to 1.00, which means that they were technically efficient, i.e., situated on the production possibility curve (see Figure 1). The other ten units obtained efficiency values of less than 1.00, which means that they were technically inefficient, i.e., their actual outputs could have been produced with less inputs. It is important to note that, due to the non-parametric estimation techniques used in DEA, it is not possible to directly test the statistical significance of the individual estimates.

Column (c), containing the inverted values of column (b), shows the potential efficiency increase. This measure is often called *Farrell-efficiency* after one of the pioneers behind the DEA-method (7). The figure 1.06 for department number 3 in column (c) means that this department could, in principle, increase its production by six percent, or produce its actual output by six percent less inputs. DEA also tells us how this could be done by a linear combination of the production technologies of two technically efficient departments or more, i.e., the "peers". The average potential efficiency increase at the units studied is estimated at seven percent in 1994.

In column (d) the departments are ranked in descending order of productivity. The ranking of the technically efficient departments, all of which have the value of 1.00 in columns (b) and (c), is calculated by comparing each of these departments with all departments but one's own. This measure is the above mentioned *super-efficiency* (8).

**Efficiency (productivity) in 1995**
The following columns give the corresponding information for 1995. Thus, column (e) shows that even in 1995, nine (but not quite the same units) of the 19 hospitals were technically efficient, i.e., obtained a score equal to 1.00.

Column (f), containing the inverted values of column (e), again shows the potential production increase. The figure 1.02 for department number 4 in column (f) means that this department could increase its efficiency by two percent. The average potential efficiency increase at the units studied is estimated at six percent in 1995.

In column (g), the departments are ranked in descending order of productivity in 1995.

**Changes between 1994 and 1995**
The last three columns reflect changes between 1994-1995 in terms of Malmquist indices. Column (h) shows the "efficiency" change component of the total productivity change, calculated as the ratio between the values in columns (e) and (b). A value above 1.00 indicates an increase in productivity, i.e., that the department has approached the best practice frontier. A value of less than 1.00 means a productivity decrease, i.e., that the department has moved away from the frontier. The departments which were technically efficient in both years, i.e., were situated on the frontier, obtain the value of 1.00 in this column, since technical efficiency has neither improved nor deteriorated.

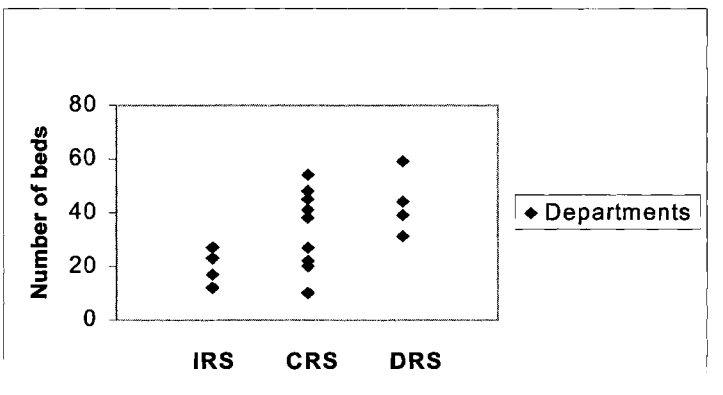
Column (i) reflects whether the best practice frontier has shifted, i.e., the "technical change" component of the overall productivity change. As shown, only two departments – 6 and 15 – have become more productive in this sense. Finally, column (j) shows the Malmquist index, which is derived by multiplying the values in columns (h) and (i). This index gives a summary measure of the productivity change, taking into account the relative change (the depart-

ment's position relative to the frontier) as well as the absolute change (the position of the frontier) (9). Four departments – 3, 6, 15, 17 – show a Malmquist index value greater than 1.

**Returns to scale**

As shown in Table 2, each year, nine of the 19 department were found to be efficient. As explained above, this means that they are operating at constant returns to scale (CRS). Among the others, four were operating at increasing returns to scale (IRS) and six (two of which are hidden in the figure) at decreasing returns to scale (DRS). Figure 2 shows how this is related to the size of the departments in terms of number of beds.

**Figure 2. Department size and returns to scale (1995).**




Obviously, departments operating at increasing returns to scale are small, while those operating at decreasing returns to scale are generally to be found among the larger ones. The picture was very similar for 1994.

**Impact of production volume (capacity)**

The results of the regression analyses to test for the impact of production volume (capacity) on efficiency are summarised in Table 3.

**Table 3. Bivariate regression results;**
**Dependent variable: Productivity (super-efficiency).**

**3a. Independent variables: Number of beds** (t-values in parentheses)

| Variable | 1994 | 1995 | Pooled |
|---|---|---|---|
| **Constant** | 1.136*** | 1.296*** | 1.214*** |
| | (9.426) | (7.733) | (12.119) |
| **Beds** | -2.82E-03 | -6.708E-03 | -4.646E-03 |
| | (-0.933) | (-1.456) | (-1.756)* |

***,**,* significant at the 1%, 5%, and 10% level, respectively.

**3b. Independent variables: Number of admissions** (t-values in parentheses)

| Variable | 1994 | 1995 | Pooled |
|---|---|---|---|
| **Constant** | 1.086*** | 1.224*** | 1.154*** |
| | (9.600) | (7.813) | (12.319) |
| **Admissions** | -1.999E-05 | -6.328E-05 | -4.019E-05 |
| | (-0.523) | (-1.083) | (-1.211) |

***,**,* significant at the 1%, 5%, and 10% level, respectively.

**3c. Independent variables: Number of deliveries** (t-values in parentheses)

| Variable | 1994 | 1995 | Pooled |
|---|---|---|---|
| **Constant** | 1.084*** | 1.227*** | 1.154*** |
| | (9.802) | (8.054) | (12.610) |
| **Deliveries** | -2.160E-05 | -7.184E-05 | -4.518E-05 |
| | (-0.514) | (-1.139) | (-1.246) |

***,**,* significant at the 1%, 5%, and 10% level, respectively.

The hypothesis of a zero effect of production volume or capacity on productivity level could not be rejected. Only one regression model, the pooled regression of productivity on number of beds, showed a significant (negative) regression coefficient. However, a closer examination of the data revealed that this was due to one outlier, department number 6, which was the most efficient according to the DEA results. Naturally, this constitutes no strong support for the null hypothesis, since the number of observations is very limited. However, the result seems to be in accordance with the findings reported in Figure 2. The efficient units were found among departments of all sizes, while the inefficient ones were found among the small and larger ones.

### Alternative model specifications

To assess the effect of different setups of the DEA model, we compare the ranking order of the hospitals (in 1994) according to the base model (3 inputs and 3 outputs), and three alternative models, where one variable or more is excluded. Alternative I excludes the output variable *other DRGs* and alternative II excludes the input variable *number of beds*. Alternative III, finally, combines the first two alternatives, thus excluding *other DRGs* as well as *number of beds*.

**Table 4. Alternative model specifications. Rank according to the base model and alternative models. Figures in bold indicate technically efficient units.**

| Hospital department | Base model (3+3) | Alt I (3+2) | Alt II (2+3) | Alt III (2+2) |
|---|---|---|---|---|
| **1** | **7** | **5** | 10 | 7 |
| **2** | **8** | **7** | 8 | 3 |
| **3** | 11 | 8 | 13 | 9 |
| **4** | 15 | 13 | 16 | 13 |
| **5** | **3** | **6** | 4 | 6 |
| **6** | **1** | **1** | **1** | **1** |
| **7** | 12 | 11 | 17 | 14 |
| **8** | **6** | 9 | **3** | 8 |
| **9** | 18 | 15 | 19 | 17 |
| **10** | 19 | 17 | 18 | 15 |
| **11** | 10 | 19 | 6 | 19 |
| **12** | **4** | **4** | 7 | 5 |
| **13** | 14 | 12 | 14 | 10 |
| **14** | 13 | 10 | 12 | 12 |
| **15** | **9** | 18 | 5 | 18 |
| **16** | 16 | 14 | 15 | 11 |
| **17** | 17 | 16 | 11 | 16 |
| **18** | **2** | **2** | **2** | **2** |
| **19** | **5** | **3** | 9 | 4 |
| **Nb of eff. Depts.** | 9 | 7 | 3 | 2 |

The row at the bottom of Table 4 shows that the number of technically efficient units decreases with the number of variables. This is due to the fact that a smaller number of variables implies a limitation of the production possibilities and thus, a lower probability for a unit being unique in some respect, i.e., to be regarded as technically efficient, because there is no unit for comparison. This also implies that no department can get a higher efficiency value as the number of variables decreases.

The alternative models imply unchanged or higher rank for most units. However, one department (number 5) obtained a lower rank in all alternatives, and five departments – 7, 8, 11, 12 and 15 – obtained a lower rank in two alternatives out of three. The fact that department 1 obtains rank number 10 instead of 7 when the variable *number of beds* is excluded (Alt II) indicates that this department utilises its beds relatively efficiently. Consequently, the hospital is "favoured" when this variable is included.

The opposite is true for e.g. department number 8, which obtains rank number 3 instead of 6 when the variable *number of beds* is excluded, indicating a relatively inefficient bed utilisation. Thus, department 8 obtains a better rank order when this variable is excluded. Accordingly, we conclude that e.g. department 19 is relatively inefficient in producing *other DRGs*, since the rank number is 3 instead of 5 when this output variable is excluded (Alt I). The department is "favoured" when the variable *other DRGs* is excluded.

# 5. DISCUSSION AND CONCLUDING REMARKS

The purpose of this study is to analyse the productivity of inpatient obstetric care in Swedish hospitals using the DEA method. An advantage of the DEA-method is that several inputs and outputs can be simultaneously included in the model. This is important since the "production technology" in health care is characterized by multiple outputs produced by multiple inputs. Our study object, inpatient obstretic care, was selected because it is possible to give a relevant and complete description of outputs as well as inputs in this area.

Data quality is generally good. By collecting the data directly from the hospital departments, we probably obtained the best possible data quality. A crucial assumption made is that the input variables *physician costs* and *costs for other staff* reflect real resource inputs. This assumption was based on the fact that in the mid-1990's, there were only very small wage differences among hospitals within the same personnel categories. Today, this is probably not the case, so that new studies should not apply that assumption, but use input measures expressed in quantities.

The results indicate a potential for enhanced productivity of inpatient obstetric care – on average seven percent in 1994 and six percent in 1995. That is, in principle, it would be possible to produce the actual output with less resources. Furthermore, the fact that DEA measures "relative" efficiency indicates that the true potential is even larger.

The volume of inpatient obstetrical care is, to a large extent, determined by demand. Since the demand for deliveries is exogenous and competition among hospitals was limited at the time, the production volume is largely determined by conditions beyond the control of the single production unit. When the number of deliveries decreases, as it did in all of the hospitals but one between 1994 and 1995, this naturally implies that output, in terms of cesarean sections and vaginal deliveries, decreases. In the short run, it may be difficult to adjust production capacity and costs accordingly. Decreasing output and constant or increasing costs by definition imply decreasing productivity.

In principle, it is not impossible to reduce costs when demand decreases, however. Personnel can be reallocated to other production lines, e.g. outpatient care. Furthermore, forecasting the demand for (inpatient) obstetric care is relatively straightforward. The adjustment to changing demand thus is a very important factor for maintaining high productivity.

Despite the fact that the average *relative* productivity increased by 1 percent from 1994 to 1995 (see Table 2, column h), the *absolute* productivity decreased in most departments. This, in turn, means (cf. Figure 1) that the production possibility curve (the best practice frontier) moved inwards - towards origo. Thus, most departments obtained a Malmquist index score less than 1, which is equivalent to a negative change in productivity.

The obstretic departments in this study are hardly representative for all obstetric departments in Sweden. Thus, it is not possible to generalize from our results. On the other hand, there is no obvious reason to believe that other hospitals are, in general, more – or less – efficient than those in the study.

DEA is an appropriate method for efficiency (productivity) analysis in health care. The method is well suited for benchmarking purposes, particularly as it indirectly shows the strengths and weaknesses of every single production unit. By including health outcome and

quality indicators among the output variables, it is, in principle, possible to further improve the analysis. This is an important area in need of further research and development. There are good reasons for adopting the DEA method in other speciality fields as well, and preferably on a larger scale, i.e., with more production units included. The more observations, the more variables can be included, for example quality indicators.

## REFERENCES

1. Charnes A, Cooper WW, Rhodes E (1978). Measuring the Efficiency of Decision making Units. European Journal of Operational Research 2:429-44.

2. Charnes A, Cooper WW, Lewin AY, Seiford LS (eds), (1995). Data Envelopment Analysis : Theory, Methodology and Applications. Amsterdam: Kluwer Academic Publishers.

3. Folland S, Goodman AC, Stano M (1993). The Economics of Health Care. New York: Macmillan.

4. Salinas-Jiménez J, Smith P (1994). Data Envelopment Analysis Applied to Quality in Primary Health Care. York: Centre for Health Economics. University of York. (Discussion Paper 124).

5. Tambour M (1997). Essays on Performance Measurement in Health Care. Stockholm: Stockholm School of Economics. Diss.

6. Eckerlund I, Ekfeldt C, Gårdmark S, Rydhström H (1996). Kvinnoklinikerna 1995. Stockholm: Spri. (Projekt-PM). (In Swedish).

7. Farrell MJ (1957). The Measurement of Productive Efficiency. Journal of the Royal Statistical Society 120. Series A, part 3:253-81.

8. Andersen P, Petersen NC (1993). Procedure for Ranking Efficient Units in Data Envelopment Analysis. Management Science 39:1261-64.

9. Färe R, Grosskopf S, Lindgren B, Roos P (1989). Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach. Carbondale: Department of Economics, Southern Illinois University. (Discussion Paper 89-3).

10. Färe R, Grosskopf S, Lovell K (1994). Production Frontiers. Cambridge University Press.

# Change-oriented patient questionnaires – testing a new method at three departments of ophthalmology

**Ingemar Eckerlund**
Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden
**Bengt Jönsson**
Professor, Health Economics, Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden
**Magnus Tambour**
Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden
**Anders H. Westlund**
Professor, Economic Statistics, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden

IV

**The use of patient questionnaires has increased widely in recent years. Their purpose, to incorporate patient perspectives into the orientation and design of health care, is, of course, commendable. However, the survey methods themselves have been less adequate, both in terms of validity and reliability, and with respect to the potential for using the results to improve the quality of health care. Presents a pilot study at three departments of ophthalmology in Sweden, involving a new method which meets reasonable demands for validity and reliability, and is explicitly change-oriented.**

## Introduction

Interest in eliciting patients' views on the quality of health care has increased substantially in recent years. This is a positive development since patients' perceptions of quality are essential for determining effectiveness – and efficiency – in health care delivery. Patient questionnaires have been the main instrument for gauging patient satisfaction, and are being used across most of the health care sector. Indisputably, these efforts have been positive in the sense that patient perspectives are receiving greater attention. This has contributed towards greater awareness of the customer perspective among health care personnel and has engaged patients in quality improvement efforts. Consequently, a range of quality shortcomings have been identified and corrected[1-4].

Attempts to survey patient perspectives have not been without problems. The survey methodology has been criticized, among other things for deficient validity and reliability, and a weak orientation towards change. Procedures for local administration of questionnaires and interpretation of responses have been questioned[5]. Considerable effort has been invested to improve the methods[3]. In this paper we describe a study involving a new method which assures a reasonable level of validity and reliability, while being explicitly change oriented. Three Swedish ophthalmology departments – at Borås Hospital, Mälar Hospital in Eskilstuna, and S:t Erik's Eye Hospital in Stockholm – participated in this pilot study. The following groups were studied: cataract patients, strabismus (squinting) patients, and diabetic patients. These segments of patients were chosen because they cover a wide range of age groups.

## Methods and data

A method called Quality, Satisfaction, Performance (QSP) was used to measure quality and focus on quality improvement. Figure 1 illustrates the QSP model that was used in the project.

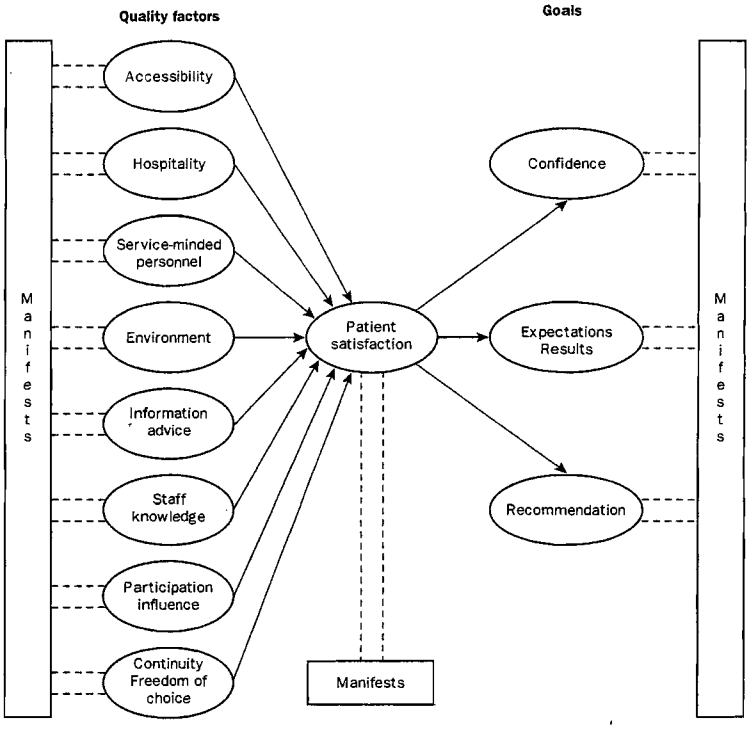
The model consists of three integrated components. One component measures the degree of patient satisfaction (and different aspects thereof) among different patient groups. Patient-perceived quality levels of various quality dimensions (quality factors) are also measured, which *a priori* are assumed to explain the variation in patient satisfaction. Each quality dimension is represented in the patient questionnaire by three to five questions, intended to represent a specific quality factor as thoroughly and reliably as possible. The questions should be specific enough to provide an operative decision-making basis for quality improvement. Finally, the model also contains a component on goals, with questions directed at what patient satisfaction should ultimately lead to, e.g. increased trust, increased likelihood for positive recommendations, etc. This aspect of the measurement model assures the validity of the model and the measurement method, since patient satisfaction is, as strongly as possible, related to the goals.

The questionnaire addressed eight different quality dimensions, using 41 questions each with a ten-grade rating scale. The eight dimensions: accessibility, hospitality, service commitment, etc., appear in the left part of Figure 1. The questions were formulated as: "What rating would you give to...?". For example, Figure 2 presents some of the questions within the "Environment" quality dimension. Each question relates to specific

International Journal of Health Care Quality Assurance
10/7 [1997] 254–259

[ISSN 0952-6862]

**Figure 1**
The QSP model




actions that a department can take if a particular dimension receives a low rating, while at the same time being important to patient satisfaction.

To measure "patient satisfaction" (PSI = patient satisfaction index) three questions were asked concerning how patients perceived the visit generally. These questions correspond to the centre part of Figure 1, and concern the degree to which patients were satisfied with the service, whether the visit met their expectations, and how closely the clinical department they visited approximated an "ideal clinical department". Based on these questions a PSI was calculated for each patient.

The last five questions in the questionnaire, which correspond to the right part in the

**Figure 2**
Examples of questions in the quality dimension "environment"

| *How do you grade* | Lowest grade | | | | | | | | | Highest grade | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Do not know |
| A – The possibility to have private conversations with reception staff | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| B – The space and comfort of the treatment room | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| C – The possibility to have private conversations with nursing staff | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

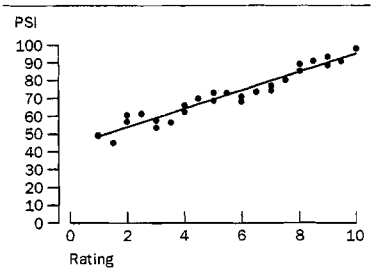
model, addressed the departmental goals. Thus, it is not enough to measure patient satisfaction and identify changes that contribute most towards increasing patient satisfaction. Rather, patient satisfaction must be defined in relation to "the goals". Measurement models and measurement methods confirm the relationship between patient satisfaction and the "goal variables". In this study we (in collaboration with the department heads) chose to ask patients questions about their expectations in terms of how the visit influenced their sight and other disease-related problems, about trust in the department, about the willingness to seek advice from departmental staff by telephone, and about whether they would recommend the department to an acquaintance with sight problems who is seeking care.

In the QSP model, the relationships between quality factors and satisfaction are expressed as equations. When limited to two variables, these equations can be illustrated graphically. Figure 3 shows the schematic relationship between a quality factor and patient satisfaction (assuming that just one quality question/manifest represents the quality factor).

Each dot in the figure represents the response from one patient, with the patient's rating of the particular quality question (manifest) plotted on the horizontal axis and the patient satisfaction index on the vertical axis. Using the QSP method, we can draw a regression line based on these data and calculate the slope of this line. The slope of the line determines the impact of the factors on PSI. The steeper the slope, the more important the factor. Therefore, the importance of a particular quality factor is determined by its impact on PSI. If the slope is steep, even a small change in the rating may substantially influence PSI.

This briefly describes the principles behind the QSP method. In practice, however, there are several manifests/questions combined to each quality factor and several quality factors. The effect on PSI is measured for each quality factor in a simultaneous estimation process. This information may then be combined with the relevant average ratings and presented in a simple four cell matrix. See Figure 4.

**Figure 3**
Relation between quality factor and PSI




In this way it is possible to identify the factors which should be prioritized to achieve greater patient satisfaction. The greatest "return" is obtained by prioritizing improvements in quality factors that land in the lower right quadrant (high impact/low rating). Factors on the left in the diagram should receive lower priority.

The associations on the right in the model (see Figure 1) are analysed simultaneously in similar fashion to show the effect that a particular increase in PSI has on the goal variables. The model is now complete, providing a method for determining the changes that would most efficiently increase the value of the goal variables.

The statistical methodology used in the measurement model is taken from modern multivariate analysis. It is based on principles applied in the partial least squares methodology (PLS)[6]. The analytical model consists of two components:

1. a structural model that describes how quality factors, patient satisfaction, and goal factors are related to each other (see Figure 1);
2. measurement models (one for each factor), that show how quality factors/patient satisfaction/goal factors are related to the survey questions, which represent the respective factors.

Somewhat more formally, the structural model may be described as follows:

$$E(\eta|\eta,\mu) = \alpha\eta + \beta\mu, \qquad (1)$$

where $\eta = (\eta_1, \ldots, \eta_p)$ is a vector of endogenous factors (here patient satisfaction and the goal factors respectively , i.e. $p = 4$), and $\mu = (\mu_1, \ldots, \mu_q)$ is a vector of exogenous factors (here quality factors to the left in the model, i.e. $q = 8$) and $E(.)$ indicates the expected value, $E(.|.)$ signifies the conditional expected value. The matrices $\alpha\,(p \times p)$ and $\beta\,(p \times q)$ contain effect parameters (from patient satisfaction to goal variables, and from quality factors to patient satisfaction).

**Figure 4**
Prioritizing matrix

| | |
|---|---|
| Low impact/high rating | High impact/high rating |
| Low impact/low rating | High impact/low rating |

The vectors $\eta$ and $\mu$ are latent, or not directly observable (i.e. measured indirectly by their respective manifests). The structural model (1) of course never includes all factors that explain the variations in patients' quality values. The model, therefore, contains a vector of errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)$, where $e = \eta - E(\eta|.)$. In the continued analysis it is assumed that $E(\eta\varepsilon) = E(\mu\varepsilon) = E(\varepsilon) = 0$.

The formal description of the measurement models is presented as follows:

$$y = A\eta + \upsilon$$
$$x = B\mu + \delta, \qquad (2)$$

where $y = (y_1, \ldots, y_m)$ is a vector of the questionnaire's manifest variables related to the endogenous factors in $\eta$ and $x = (x_1, \ldots, x_n)$ is a vector of the questionnaire's manifest variables related to the exogenous factors in $\mu$ (i.e. $m + n$ = number of questions in the questionnaire).

The matrices $A$ $(m \times p)$ and $B$ $(n \times q)$ show how the manifests correlate to their respective latent factors. In the statistical analysis, the impact parameters ($\alpha$ and $\beta$) are determined, and the rating associated with the latent variables $\eta$ and $\mu$ – in principal $E(\eta)$ and $E(\mu)$ – are estimated. Furthermore, $A$ and $B$ are calculated to provide information concerning how the manifests correlate. The statistical method is iterative, having the characteristic that it converges towards statistical estimates which are consistent and have the best possible precision[7].

### Data

The data was collected via a questionnaire survey at three ophthalmology departments during May and June of 1994. Patients in the three "segments" – cataracts, strabismus, and diabetes – answered (partly different) questionnaires containing 49 questions about quality at the departments. The questionnaires were distributed to 150-200 consecutive patients in each segment at the different departments. This was done in conjunction with an ambulatory visit, and patients were requested (under normal anonymity protection) to fill out and return the questionnaire to the department within one week. Approximately 1,500 patients were invited to participate in the study. Written reminders, and a new questionnaire, were sent to patients who did not respond within ten days. The final response rate averaged 80 per cent.

## Results

Table I shows the patient satisfaction index (PSI) at each department, and for each patient segment. The maximum index value is 100.

**Table I**
Patient satisfaction index (PSI) for different patient groups and departments (maximum value = 100)

| | Diabetic | Cataract | Strabismus | All |
|---|---|---|---|---|
| **Borås** | 85.7 | 90.5 | 81.7 | 86.2 |
| **Eskilstuna** | 87.6 | 92.4 | 89.2 | 89.7 |
| **S:t Erik** | 83.0 | 95.6 | 85.3 | 87.3 |
| **All** | 85.4 | 92.3 | 85.5 | 87.8 |

The table shows that PSI is consistently high. All departments received an index value between 86 and 90, which is very good. Other studies based on the QSP model often reflect substantially lower index values. In comparison, the PSI for community health centres is estimated to be 68, according to the Swedish Customer Satisfaction Barometer[8].

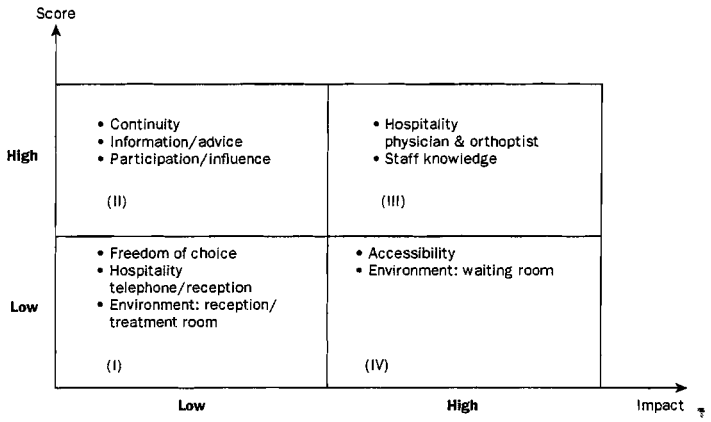
Although the index figures are consistently high, satisfaction varies somewhat among the different patient segments and departments. For example, cataract patients were found to be the most satisfied patient group at all three departments. The lowest PSI for cataract patients (90.5) was higher than the highest index figure for the other groups.

In addition to patients rating the different quality dimensions, the QSP method also shows how these dimensions impact on PSI. Figure 5 illustrates the rating and impact of the quality dimensions. The results of the questionnaire from strabismus patients at the department of ophthalmology in Eskilstuna are used as an example.

The lower left field (I) presents the quality dimensions that received relatively low ratings and have little impact on PSI. The upper left field (II) shows quality dimensions which received high ratings but also have little impact on PSI. Field III reflects dimensions which the departments were considered to be good at (high rating) and which also have a major impact on PSI. Finally, field IV shows the dimensions which patients view to be important, but which received relatively low ratings. As can be noted, the dimensions in Figure 5 are not exactly the same as in Figure 1. This is due to a respecification of the quality factors as a result of the estimation process.

The different fields provide departure points for choosing optimum treatment strategies. Field IV contains dimensions that the department should focus on in their quality improvement efforts, e.g. to improve accessibility (better telephone service, reduced waiting times, etc.) and the waiting room environment. Field III contains dimensions for which the department should maintain the level of quality, since even these greatly influence patient satisfaction. Therefore, the department should

**Figure 5**
Example of prioritizing matrix: strabismus Eskilstuna




attempt to maintain the level of the service commitment by physicians and orthoptist. Field II contains dimensions where the department may potentially reduce its efforts. The resources for, e.g. information and advice may possibly be reduced without jeopardizing the level of patient satisfaction. Finally, field I includes dimensions that should not be prioritized in terms of increasing patients' satisfaction. Here, this could be interpreted to mean that the department in Eskilstuna need not necessarily commit resources towards increasing freedom of choice or improving the reception area/treatment room environment.

Similar diagrams for the other departments and patient groups show similar results. For example, all patient groups in the three departments gave low ratings to accessibility and freedom of choice. However, while accessibility plays a major role in satisfaction in eight of nine patient groups, freedom of choice does not appear to carry the same importance. Consistently, patients gave high ratings to the knowledge of physicians and other staff, including their service commitment.

Finally, with regard to the right side of the measurement and analysis model (Figure 1), the results show that, for diabetes and strabismus patients, PSI had the strongest (positive) impact on confidence in the department. This suggests that if the departments, through various improvement strategies for these patients, are able to achieve higher PSI, this would increase both confidence in the department and the tendency for patients to recommend the department to others with sight problems. For cataract patients, PSI had the strongest impact on expectations of how the visit would influence sight and problems related to the disease. In this patient group, increased PSI is mainly associated with high expectations of the results of treatment.

## Discussion and concluding remarks

What distinguishes this model from most others used in health care is that it not only measures the degree of satisfaction but also the impact that various quality dimensions (factors) have on satisfaction. Consequently, both patients' quality opinions and their quality dimension preferences are identified.

Since experience shows it is difficult for patients to rank different quality dimensions in a questionnaire, the basis for prioritization must be derived from the measurement model. The method does this, which is a major advantage. Thus, the method provides a reliable base, grounded in patients' own values, for prioritizing among different measures to improve patient-perceived quality.

Another advantage with the QSP method is the linkage to the "goal side", which secures validity in the model. This aspect of the model requires the user to specify organizational goals. A department obviously cannot aim "only" towards satisfying patients. Other longer-term goals are also important to an organization's future. If a department is to survive and develop, it must define a clear and comprehensive quality strategy, of which the short-term goals are a part.

The study presented here has several characteristics of a pilot study, and therefore does not claim to have applied the methodology in

its final form. Naturally, the methodology for quality improvement must also be subject to quality improvement processes. For example, the formulation of survey questions is an important area that requires further research.

Regarding the results of the study, we shall not go into far-reaching interpretations here. An obvious result, however, is that cataract patients were clearly more satisfied than the other patient groups. This corresponds well to the experiences reported by the staff. The explanation appears to be that surgical patients are involved, and since cataract surgery is usually quite successful it leads to radical improvement in patients' quality of life. This is not the case with strabismus and diabetes, where patients are usually subjected to long treatment series with less obvious results. Even the age factor may play a role. Cataract patients are usually older, and experience shows that older patients report higher satisfaction[3].

Distinct differences were also found among the different groups with regard to individual quality dimensions. "Freedom of choice", i.e. the opportunity for patients to choose their appointment time and physician (or other caregiver) generally received a relatively low rating from patients. It can be noted, however, that freedom of choice does not greatly influence PSI.

Further analysis is probably most appropriately done locally at the participating departments themselves. It should be noted that a health economics' perspective is a prerequisite, i.e. decisions must be preceded by thorough assessment of the costs and effects associated with the various changes.

Finally, it can be noted that the project has generated major interest both among other ophthalmology departments and other areas of medicine. We believe that the method could play an important role in developing systems for continuous quality improvement in health care. Follow-up and assessment of "patient satisfaction" is critical for achieving efficiency in health services.

## References

1 Cleary, P.D. and McNeil, B.J., "Patient satisfaction as an indicator of quality care", *Inquiry*, Vol. 25, 1988, pp. 25-36.
2 Lewis, J.R., "Patient views on quality care in general practice: literature review", *Social Science and Medicine*, Vol. 39, 1994, pp. 655-70.
3 Zastowny, T.R. *et al.*, "Patient satisfaction and experience with health services and quality of care", *Quality Management in Health Care*, Vol. 3, 1995, pp. 50-61.
4 "The challenge of quality", *Inquiry* (special issue), Vol. 25, 1988.
5 Ehnfors, M. and Smedby, B., "Patient satisfaction surveys subsequent to hospital care: problems of sampling, non-response and other losses", *Quality Assurance in Health Care*, Vol. 5, 1993, pp. 19-23.
6 Dijkstra, T.K., "Some comments on maximum likelihood and partial least squares methods", *Journal of Econometrics*, Vol. 22, 1983, pp. 67-90.
7 Fornell, C. and Cha, J., "Partial least squares", in Bagozzi, P. (Ed.), *Advanced Methods of Marketing Research*, Blackwell, 1995, pp. 52-78.
8 Fornell, C., "A national customer satisfaction barometer: the Swedish experience", *Journal of Marketing*, Vol. 56, 1992, pp. 6-21.

# Patient satisfaction and priority setting - an economic approach

Ingemar Eckerlund, Jan A Eklöf, Jörgen Nathorst-Böös

V

**Abstract**

In recent years, various methods for measuring patient satisfaction have been applied as part of quality improvement programmes. However, the validity and reliability as well as the applicability and change-orientation of adopted methods have been questioned. Furthermore, most methods pay no specific attention to economic aspects. The purpose of this study is to analyse if and how priorities are influenced when an economic perspective is explicitly included. Data were compiled by a patient survey at a gynaecology clinic, inquiring about the patients' views on various quality dimensions, and their willingness to pay for proposed improvements. The parameters of the model are estimated with simultaneous equation methods, based on partial least squares technique. We compare the ranking of proposed improvements derived from a patient satisfaction index, a cost-effectiveness analysis and a cost-benefit analysis, respectively. Our results show that even a methodologically appropriate measurement of patient satisfaction may lead to a cost-ineffective priority setting unless economic consequences are explicitly taken into consideration. Further, it is demonstrated how an analysis including patient preferences as well as economic aspects may be carried out.

# 1. INTRODUCTION

Patient preferences measured in terms of satisfaction indicators, are being increasingly recognised as an important basis for priority setting in the health services and for improving the quality of care (1, 2, 9). The most common method for systematic assessment of patient satisfaction in health care is patient surveys (questionnaires). In Sweden, the first generation of patient questionnaires was introduced in the mid-1980's (13). They helped to focus on the patient perspective, but had certain deficiencies both in terms of validity and reliability, and with regard to their potential as a basis for quality improvements. However, several new methods were launched in the 1990's, claiming to meet demands for statistical quality as well as change-orientation and relevance better than earlier methods (4, 7, 10, 12, 14). Still, economic aspects are seldom included, i.e., measures identified as effective for improving patient satisfaction are usually not evaluated with regard to their costs and benefits.

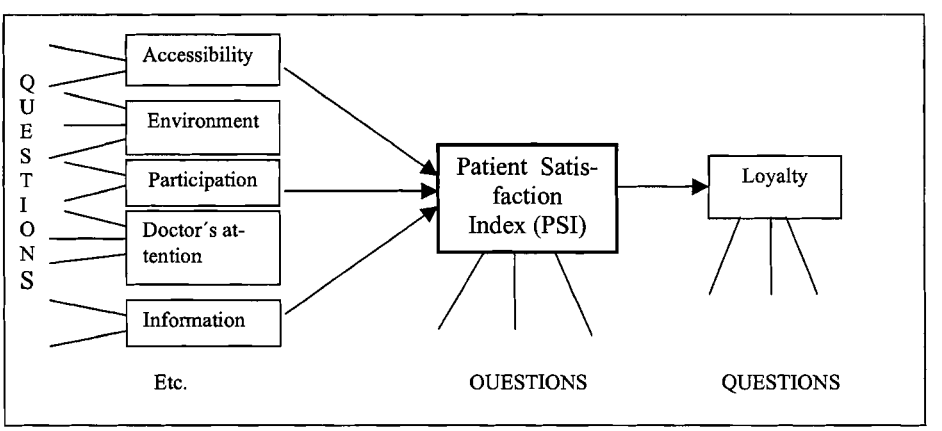
The purpose of this study is to analyse if and how the explicit inclusion of an economic perspective influences priorities based on patient views. This is accomplished in the following sequence of steps presented in the paper after the initial part devoted to genuine patient satisfaction measurements: i) a cost-effectiveness analysis of certain proposed changes/ improvements, ii) a cost-benefit analysis based on the patients' willingness-to-pay for these changes, and iii) a correlation analysis of priorities based on patient satisfaction, willingness to pay, cost-effectiveness and cost-benefit analysis, respectively.

The following presentation starts with a description of the theoretical model for patient satisfaction measurements and an outline of the empirical approach in Section 2. The results of the patient satisfaction study, the cost-effectiveness calculations and the cost-benefit analysis are presented in Section 3. The paper ends with a discussion and some concluding remarks in Section 4.

# 2. METHODS AND MATERIAL

The method applied for quality assessment based on a patient satisfaction approach, QSP (Quality, Satisfaction, Performance), is illustrated in Figure 1.

**Figure 1. The QSP-model**



The QSP model consists of three integrated components. One component measures the degree of patient satisfaction (PSI), usually through three questions. Patient-perceived quality levels of various quality dimensions (quality factors) are also measured, which are, a priori, assumed to explain the variation in patient satisfaction. Each quality dimension is represented by three to six questions in the patient questionnaire, intended to represent a specific quality factor as thoroughly and reliably as possible. The questions should be specific enough to provide an operative decision-making basis for quality improvement. Finally, the model also contains a component on goals, with questions directed at what patient satisfaction should ultimately lead to, e.g., increased trust, increased likelihood of positive recommendation and related loyalty indicators. This aspect of the measurement model links patient satisfaction to the goals of the health care provider.

The statistical methodology used in the measurement model is based on a state-of-the-art multivariate analysis based on principles applied in the partial least squares methodology, PLS (3, 8). The analytical model consists of two components:

1) a structural model describing how quality factors, patient satisfaction, and goal factors are related (see Figure 1);
2) measurement models (one for each factor), that show how quality factors, patient satisfaction and goal factors are related to the survey questions, which represent the respective factors.

More formally, the structural model may be described as follows:

$$E(\eta|\eta,\mu) = \alpha\eta + \beta\mu, \qquad (1)$$

where $\eta = (\eta_1,\ldots,\eta_p)$ is a vector of endogenous factors (here patient satisfaction and the goal factors respectively), and $\mu = (\mu_1,\ldots,\mu_q)$ is a vector of exogenous factors (here quality factors to the left in the model). E (.) indicates the expected value, and E (.|.) signifies the conditional expected value. Matrices $\alpha\ (p \times p)$ and $\beta\ (p \times q)$ contain effect parameters, from patient satisfaction to goal variables, and from quality factors to patient satisfaction.

Vectors $\eta$ and $\mu$ are latent, or not directly observable (i.e., measured indirectly by their respective manifests). The structural model (1) naturally never includes all factors explaining the variations in patients' quality values. The model therefore contains a vector of errors $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_p)$, where $\varepsilon = \eta - E(\eta|.)$. In the continued analysis, it is assumed that $E(\eta\varepsilon) = E(\mu\varepsilon) = E(\varepsilon) = 0$.

The formal description of the measurement models is presented as follows:

$$\begin{aligned} y &= A\eta + \upsilon \\ x &= B\mu + \delta, \end{aligned} \qquad (2)$$

where $y = (y_1,\ldots,y_m)$ is a vector of the questionnaire's manifest variables related to the endogenous factors in $\eta$, and $x = (x_1,\ldots,x_n)$ is a vector of the questionnaire's manifest

variables related to the exogenous factors in $\mu$ (i.e., m + n = number of questions in the questionnaire).

Matrices *A* (m × p) and *B* (n × q) show how the manifests correlate with their respective latent factors. In the statistical analysis, the impact parameters ($\alpha$ and $\beta$) are determined, and the rating associated with the latent variables $\eta$ and $\mu$ - in principal $E(\eta)$ and $E(\mu)$ - is estimated. Furthermore, *A* and *B* are calculated to provide information on how the manifests correlate. The statistical method is iterative with the characteristic of converging toward statistical estimates that are consistent and have the best possible precision (8).

In earlier studies in the health services, the QSP method has been shown to produce relevant results (4, 6, 7). Especially, it identifies the most effective measures of improving PSI. A weakness in common with other known methods is the lack of an economic dimension. The analyses of the patients' valuation of various quality factors and proposed improvements do not take into account the cost-effectiveness of different improvement measures. For this purpose, we must estimate the "PSI production function", which is done in the cost-effectiveness analysis. Furthermore, to know how much should be spent on improvements, we must also do a cost-benefit analysis.

The current study is based on an earlier study of patient satisfaction at the gynaecology clinic at the Karolinska Hospital in Stockholm, Sweden, carried out in 1997 with the QSP method (7). In that study, three main areas calling for improvement were identified, namely *accessibility*, *environment* and *participation*. Now, these areas are analysed in further detail in terms of current patient satisfaction. This follow-up study is combined with a study of patient valuation – regarding satisfaction and willingness to pay, respectively – of possible improvements and relative to the calculated costs of these improvements.

Thus, we use a reduced model including three of the seven quality dimensions in the above mentioned QSP-study. We single out the three dimensions (latent variables) which were found to be important in determining the patient satisfaction level and, at the same time, being among the least appreciated by the patients (high impact *and* low scores). Each of the latent variables – *accessibility, environment* and *participation* – is measured with four questions, i.e., manifest variables[1].

To obtain an assessment of overall patient satisfaction, we asked three questions about the visit as a whole; to what extent the patient was satisfied with the service, if the visit fulfilled the expectations and how the actual clinic compared to an ideal clinic. Finally, three questions were asked concerning the patients' loyalty to the clinic; the trust in the clinic, whether the patient would come back for another visit, and if she would recommend a friend to visit the clinic.

Thus, the study consists of three parts. This *first part* constitutes a replication of the above-mentioned study conducted in 1997. It should be noted that the replication was done in a most condensed form as the original study included seven different latent variables. In addition to the variables studied here, the four latent variables medical care, doctor's attention, nurse's attention and information were included in the initial study. All of these received very high scores and/or low impact levels, and were thus considered not to be so important for focused

[1] In the WTP-study, two of the manifest questions – relating to waiting-room functionality – were pooled together.

improvement work. However, the relationship between costs and patient valuation was not explicitly considered for them.

In the *second part*, we included questions concerning possible improvements expressed in certain detail, and asked the respondents to indicate on the same scale what score they would give if such a change was implemented. In this way, we obtain an estimate of how much such a change is valued in terms of improved patient satisfaction. The following proposed modifications have been modelled[2]:

***Accessibility***
- Extended telephone hours by two hours per working day (*telephone hours*);
- "Guaranteed" answer within three telephone signals (*telephone answer*);
- The doctor calls back the same day as the initial contact (*call back*);
- Open clinic two hours every Wednesday evening (*opening hours*).

***Environment***
- Improved sign-posts guiding the visitor to the clinic (*sign-posts*);
- TV, coffee-machine and new magazines in the waiting room (*waiting-room*);
- Offering lockable wardrobe in the proximity to the clinic (*wardrobe*).

***Participation***
- The patient is always offered to choose the appointment time for her next visit (*next visit*);
- The patient is routinely offered to see the same doctor every time (*same doctor);*
- The scheduled visit time is extended by 10 minutes (*visit length*);
- The patient is asked in advance whether students may be present (*medical students*).

In the *third part*, the respondents were asked to indicate whether they would be willing to pay a certain amount (on top of the current patient charge) to obtain the indicated improved service/treatment. Eight different payment alternatives ranging from SEK 10-150 were defined and randomly given to an equal number of the patients.

**Implementation of the study**

The study was conducted during the period August 1998 – February 1999 at the gynaecology clinic at Karolinska hospital, Stockholm. All patients visiting the clinic in this period (the first time during the period they visited the clinic) received a survey questionnaire to be filled in either directly after the visit, or later and if so, to be returned by mail to the clinic. The survey questionnaire included a total of 43 closed response questions and two open questions (where the respondents could indicate any positive and/or negative aspects relating to the clinic). By and large, the same manifest questions as in 1997 were asked but a number of these were rephrased. The questionnaire design was pre-tested on both employees of the hospital, and actual patients. The structure of the questionnaire, which took on average some 15 - 20 minutes to fill in, is shown in the Appendix.

In total, 800 original questionnaires were distributed to patients in the period, coinciding with the number of first time visits in the study period. About 100 patients received a reminder including a new questionnaire and this procedure gave a gross response rate of 88 percent. After discarding forms with less than 70 percent filled-in answers (or inconsistencies in answers

[2] The proposed changes were based on a review of the answers of the patients in the 1997 study when asked to report on "This is what I am especially dissatisfied with", and also on discussions with the clinic staff.

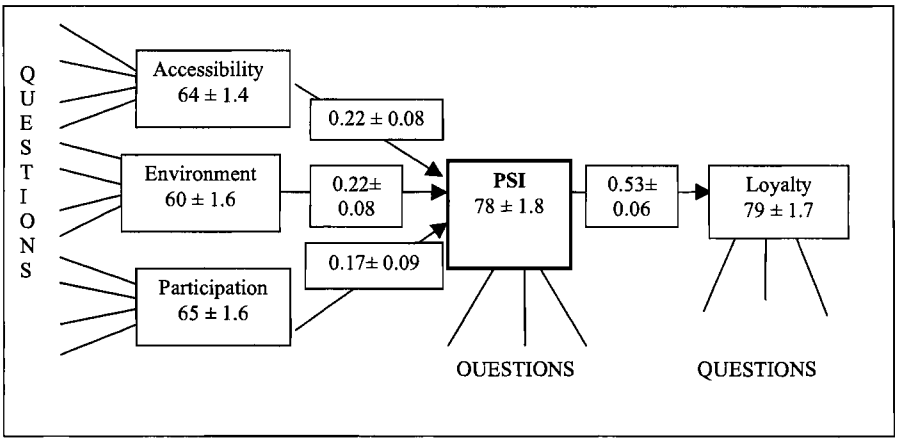
according to the test program), 657 forms (82 percent) were accepted and thus included in the analysis.

## 3. RESULTS

**Patient satisfaction**

The estimated condensed QSP-model is summarised in Figure 2. Calculated average scores and impact coefficients are given together with 95 percent confidence intervals (CI).

**Figure 2. The QSP-model – the patients' valuation of actual service**




The scores are reported on the scale 0 - 100, where 0 means totally dissatisfied and 100 totally satisfied. This transformation of the original 10-grade measurement scale is performed to make the presentation as useful as possible also for policy purposes. The Patient Satisfaction Index (PSI) was 78 (76.2-79.8; 95% CI). The "loyalty" score was 79. The three quality factors – accessibility, environment and participation – received relatively low scores, 60-65. The impact coefficients, e.g. 0.22 for accessibility, should be interpreted as the estimated effect on PSI of a one-unit change in the score for quality factors. Correspondingly, the figure 0.53 to the right tells the effect on "loyalty" of a one-unit change in PSI. With one exception, the estimated scores and impact coefficients were lower than in the 1997 study, mainly due to the actual model being partial (only three out of the originally seven dimensions, latent variables, were included here). Only accessibility received a higher score than in the earlier study. Still, the statistical quality of the model is generally acceptable (as indicated by the confidence intervals in Figure 2).

Table 1 shows current scores, relative weights for the various questions and estimated impact on PSI, and also the scores if the proposed improvements are implemented.

**Table 1. Patient satisfaction – Base case and change results**

| Factor | Score actual | Relative Weight | Impact on PSI | Score after change |
|---|---|---|---|---|
| | | | | |
| **PSI** | **78** | | | **92** |
| **Loyalty** | **79** | | | **87** |
| | | | | |
| **Accessibility** | **64** | | **0.22** | **89** |
| - telephone hours | 68 | 0.31 | 0.161 | 89 |
| - telephone answer | 56 | 0.22 | 0.116 | 89 |
| - call back | 49 | 0.25 | 0.127 | 89 |
| - opening hours | 70 | 0.22 | 0.114 | 89 |
| | | | | |
| **Environment** | **60** | | **0.22** | **84** |
| - sign-posts | 68 | 0.32 | 0.161 | 88 |
| - waiting-room comfort | 60 | 0.36 | 0.178 | (82) |
| - waiting-room convenience | 60 | 0.24 | 0.120 | (82) |
| - wardrobe | 38 | 0.07 | 0.037 | 86 |
| | | | | |
| **Participation** | **65** | | **0.17** | **89** |
| - next visit | 58 | 0.22 | 0.098 | 89 |
| - same doctor | 56 | 0.29 | 0.124 | 94 |
| - visit length | 69 | 0.33 | 0.140 | 84 |
| - medical students | 69 | 0.16 | 0.069 | 88 |

It is seen that the highest actual scores (where the patients are most satisfied) are obtained for opening hours, visit length and the presence of medical students, while the lowest scores are noted for wardrobe and (doctor's) call back. The relative weights (standardised to 1 for each latent variable) are fairly equal for the majority of the manifests, while a few are significantly different (wardrobe and medical students are lower than the average; and telephone hours, sign-posts, waiting-room comfort as well as visit length are higher than the average).

The column "impact on PSI" indicates how much a one-unit increase in the score for the respective manifest variable would increase the PSI-score (based on using the absolute – not relative – weight of each manifest variable). From the above table, it is seen that the factor relating to waiting-room comfort has the strongest single effect (with telephone hours and sign-posts in shared second place).

The effects of all the proposed improvements are shown in the last column of Table 1. Using the same impact coefficients and relative weights, the figures tell us what would happen if all the proposed improvements were implemented at the same time. It is seen that the PSI would increase by 14 units (from 78 to 92) and loyalty by 8 units (from 79 to 87). It should be noted that the questions in the survey were spelled out in such a way that the respondents considered one proposed improvement at a time. As it would be realistic to assume that the proposals are not independent, the presented results should be seen as a simplification of the real effects. Possible interactions between packages of improvements have not been considered so far.

As shown in Table 1, all latent and manifest variables are increased as a consequence of the proposed improvements. This means that the patients find them all to be positive (representing real improvements). This is true also for almost each individual respondent answer (hardly any "inconsistencies" were found in the database is this respect).

The pending question now is whether the indicated improvements are really worth implementing. In the next section, a cost-effectiveness analysis is reported, specifying the relative value of improvements in each manifest variable (expressed in terms of PSI), and comparing these with the associated costs.

**Cost-Effectiveness Analysis**

What distinguishes the QSP-model from most other methods used for quality measurement in health care is that it not only measures the degree of patient satisfaction but also the impact of various quality dimensions (factors) on satisfaction. Consequently, one of the model's strengths is that the patients' quality opinions as well as their quality dimension preferences are identified. Thus, the identified opportunities for change are based on actual patients' opinions.

However, the model does not account for the cost-effectiveness of the potential changes. In order to remedy this inherent weakness, we estimated the costs as well as the effects (impact on Patient Satisfaction Index, PSI) of various specified potential changes in the three main areas (domains) considered – accessibility, environment and participation. The costs – mainly in terms of costs for additional personnel needed for extended telephone hours, etc, but also for waiting-room equipment, and so on – were estimated in collaboration with the clinic's financial controller and only costs at the clinic level are included. The cost-effectiveness ratios in Table 2 were calculated as the monetary cost to achieve a one-unit increase in PSI from the respective manifest variable.

**Table 2. Cost-effectiveness ratios for proposed changes**

| Change | Cost per visit (SEK) | PSI-change[3] | Cost-effectiveness ratio[4] |
|---|---|---|---|
| | | | |
| *Accessibility* | (88) | | |
| - telephone hours | 20 | 3.381 | 5.92 |
| - telephone answer | 66 | 3.828 | 17.24 |
| - call back | 1 | 5.080 | 0.20 |
| - opening hours | 1 | 2.166 | 0.46 |
| | | | |
| *Environment* | (10) | | |
| - sign-posts | 1 | 3.220 | 0.31 |
| - waiting-room | 8 | 3.278 | 2.44 |
| - wardrobe | 1 | 1.776 | 0.56 |
| | | | |
| *Participation* | (315) | | |
| - next visit | 100 | 3.038 | 32.92 |
| - same doctor | 100 | 4.712 | 21.22 |
| - visit length | 110 | 2.100 | 52.38 |
| - medical students | 5 | 1.311 | 3.81 |

It is noted that the cost-effectiveness ratios are varying within very wide margins, which is mainly due to large differences in cost between the various improvement measures. The lowest cost-effectiveness ratio (the cheapest way of increasing PSI) has been calculated for the

[3] The PSI-change shows how much PSI will increase as a consequence of the increased scores (Table 2).

[4] The cost-effectiveness ratio is defined as the calculated cost per unit increase in PSI, i.e., the cost per visit divided by the PSI-change

improvement consisting of the doctor calling back the same day the patient has called. The highest cost-effectiveness ratio, SEK 52.38, stands for extended length of visits.

The remaining question is now whether these improvements are worth the additional costs (expressed in monetary terms) in the eyes of the actual patients. For that purpose, it is necessary to compare the costs and benefits measured in a single unit, which is why the additional willingness to pay study was included.

**Cost-Benefit Analysis**

This part of the study was carried out using the standard methodology for measuring willingness-to-pay that has been developed and applied in various societal contexts in recent years (11). Patients were asked whether they were willing to pay an additional X SEK per visit (on top of the current patient charge, which amounts to 120 SEK), provided that the respective improvements were implemented. The patients were randomly distributed into eight groups of equal size that received different bids (X), from an additional 10-150 SEK (1 US$=10.50 SEK) per visit and improvement. Thus, 100 patients were allocated to each pay alternative.

The result of the WTP-study differs between components. The mean willingness to pay was calculated for each of the considered alternatives. A logistic as well as a probit model have been estimated. The results of the willingness-to-pay study are summarised in Table 3, which shows the average willingness-to-pay in SEK for the various improvements. It is based on the Logit model estimation[5].

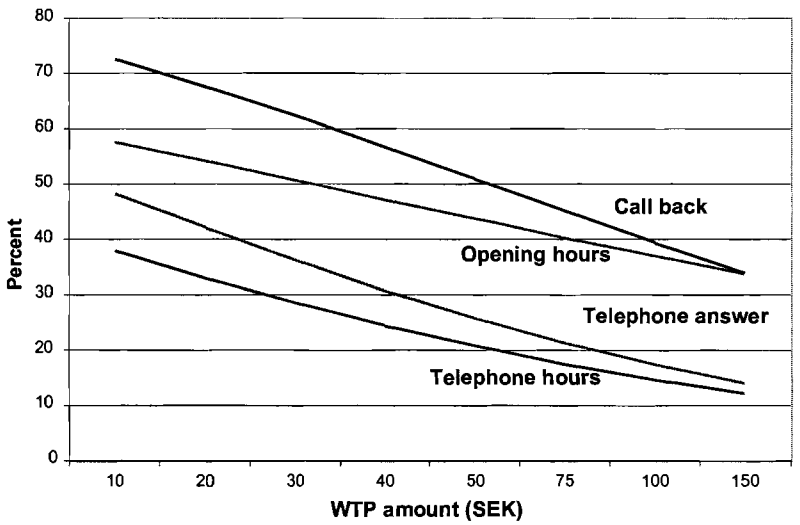
**Table 3. Willingness-to-pay (WTP) for various changes (based on the Logit model)**

| Change | Mean WTP (SEK) |
|---|---|
| | |
| *Accessibility* | |
| - telephone time | 10.68 |
| - telephone hours | 13.01 |
| - call back | 26.52 |
| - opening hours | 23.83 |
| | |
| *Environment* | |
| - sign-posts | 3.52 |
| - waiting room | 7.69 |
| - wardrobes | 8.31 |
| | |
| *Participation* | |
| - next visit | 14.31 |
| - same doctor | 31.12 |
| - visit length | 12.73 |
| - medical students | 10.78 |

[5] The results are rather similar also in the Probit model case. However, due to the fact that the distribution between alternatives is not normal, we decided to use the Logit alternative in the further analysis.

Figure 3 shows the estimated percentage of the patients (based on the Logit model) willing to pay an additional X SEK per visit for the respective alternatives associated with the ***accessibility*** latent variable. It is shown that a relatively smooth downward function is estimated for each of the four considered improvement situations.

**Figure 3. Distribution of WTP (percentage answering yes); Logit model**




A comparison between the costs of realising the proposed improvements and the patients' valuation (willingness-to-pay) demonstrates that willingness-to-pay is higher for some improvements while for some, they are lower than the cost. Table 4 shows the ratio between the average (mean) willingness-to-pay (Table 3) and the cost per visit (Table 2) for the various improvements.

The effect on patient satisfaction (PSI) of running the improvements with a WTP-cost ratio equal to or above 1 will be an increase by four units (from 78 to 82). At the same time, the loyalty index would increase by three units (from 79 to 82). The cost increase per visit for this improvement package is estimated at 9 SEK[6].

[6] Based on an estimated 4 200 visits per year, the total costs would be 37 800 SEK.

**Table 4. Willingness-to-pay (WTP)-cost ratios for various changes**

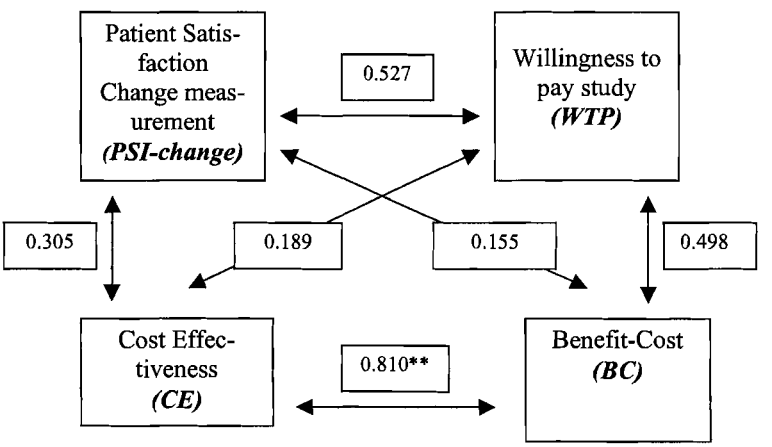
| Changes | WTP-cost ratios |
|---|---|
| | |
| *Accessibility* | |
| - telephone hours | 0.54 |
| - telephone answer | 0.20 |
| - call back | 26.52 |
| - opening hours | 23.83 |
| | |
| *Environment* | |
| - sign-posts | 3.52 |
| - waiting-room | 0.96 |
| - wardrobe | 8.31 |
| | |
| *Participation* | |
| - next visit | 0.14 |
| - same doctor | 0.31 |
| - visit length | 0.12 |
| - medical students | 2.16 |

**Using the alternative approaches in priority setting**

To summarise our findings on whether and how the priorities based merely on patient preferences are influenced by the inclusion of an economic perspective, we finally compare *the rank order of the various improvements* according to the alternative priority rules, i.e., *PSI-change, willingness-to pay, cost-effectiveness* and *benefit-cost ratio.* We find the following Pearson correlation coefficients:

**Figure 4. Correlation between the study approaches**

** Correlation significant at the 0.01 level (2-tailed)




The low correlation coefficients between the PSI-change approach and those based on cost-effectiveness (CE) and benefit-cost ratio (BC) clearly demonstrates that the ranking of the

various improvements, i.e., the priority setting, is influenced by the inclusion of an economic dimension. The correlation between PSI-change and WTP is also relatively low.

Thus, carrying out a traditional patient satisfaction analysis and using those results directly for priority setting is not enough. To appreciate the patients' valuation of possible improvements, either a PSI-change study or a WTP study is needed, both of which enable us to rank the alternative improvements. However, to reach cost-effective decisions, we must estimate the "PSI production function", which is done in the cost-effectiveness analysis. To find out whether various improvements are worth their costs, we also need a WTP-approach.

## 4. DISCUSSION AND CONCLUDING REMARKS

Patients' perceptions of quality are essential for determining effectiveness and efficiency in health care delivery, and should thus be utilised as a basis for decisions on quality improvements. Even if this is done to an increasing extent, economic aspects are seldom included in a proper way. One of the advantages of the QSP-model is its focus on change. Several studies have shown that the QSP approach can produce very useful and policy relevant results. However, from a health economics perspective, the lack of economic content in the basic model is a serious deficiency. For instance, this means that in rating the various quality factors, patients do not take cost-effectiveness into account. This, in turn, means that the priorities derived by the model may be cost-*in*effective (5).

This study is an attempt to remedy that deficiency by estimating the costs, as well as the effects, of various improvements. In that way, we are able to rank the potential improvements according to their cost-effectiveness ratios. In addition we also asked the patients to evaluate the benefits in terms of their willingness to pay for the changes, in order to rank the changes according to their benefit-cost ratios.

As shown, the ranking between various improvements is strongly influenced by the inclusion of an economic dimension. This points at the importance of looking at both the benefit and the cost side, when considering alternatives as concerns changes. We also found disagreement between cost-effectiveness and cost-benefit ranking in a few instances. This indicates a need for further analysis and model development.

The improvement *call back* gets the highest rank except for the WTP-ranking. Otherwise, the ranking differ much between the ranking order lacking an economic dimension and those based on cost-effectiveness and benefit (WTP)-cost ratio, respectively. The second and third priorities according to "PSI-change" (*same doctor* and *telephone answer*) are relatively cost-ineffective, i.e., they have relatively high cost-effectiveness ratios. Moreover, they have WTP-cost ratios of less than 1, which means that the patients are not willing to pay the costs of the changes.

Assuming a decision rule that defines a cut-off value for the WTP-cost ratio of at least 1 means that those improvement measures, which cost less than or equal to what the patients are willing to pay, shall be implemented. In our study this would mean that the following improvements – and only these – should be undertaken.

- The doctor calls back the same day as the initial contact (*call back*)
- Open clinic two hours every Wednesday evening (*opening hours*)

- Offering lockable wardrobe in the proximity to the clinic (*wardrobe*)
- Improved sign-posts guiding the visitor to the clinic (*sign-posts*)
- The patient is asked in advance whether students may be present (*medical students*)

We conclude that even if patient views are very important as a basis for improvement decisions, it is necessary to supplement it with economic analysis in order to avoid "wrong" decisions, i.e., sub-optimisation. Naturally, it is important to try to meet patient demand for continuity, i.e., seeing the same doctor on a follow-up visit. But it is not enough in terms of a basis for priority setting. There are other improvements that – according to patient views – are more important when economic aspects are considered, and thus should be given higher priority.

To sum up, our findings suggest that it is possible to expand the QSP model with an economic dimension. Thereby, one can get a clear picture of the costs and effects of various improvement measures – and a relevant basis for continuous, cost-effective quality improvement.

## REFERENCES


1. Cleary, P, McNeil, B. Patient satisfaction as an indicator of quality of care. Inquiry1988;25:25-36.

2. Coulter, A. Paternalism or partnership? BMJ 1999;319:719.20.

3. Dijkstra, TK. "Some comments on maximum likelihood and partial least squares methods". Journal of Econometrics 1983;22:67-90.

4. Eckerlund, I, Jönsson, B, Tambour, M, Westlund, A. Change-oriented patient questionnaires : testing a new method at three departments of ophthalmology. International Journal of Health Care Quality Assurance 1997;10:254-259.

5. Eckerlund, I, Eklöf, J, Nathorst-Böös, J. Patient satisfaction and priority setting in ambulatory health care. Journal of Total Quality Management 2000;7:S967-S978

6. Eklöf, J, Westlund, A. User satisfaction in the public sector : level and trends based on the Swedish Customer Satisfaction Index. Stockholm School of Economics. Research report (in press).

7. Ekfeldt-Sandberg, C., et al. Hur kan vi mäta patienternas tillfredsställelse med vården? Jämförande studie av metoderna Quality Satisfaction Performance, QSP och Kvalitet Ur Patientens Perspektiv, KUPP. Stockholm : Spri, 1999. (Spri rapport 492). (In Swedish)

8. Fornell, C, Cha, J. Partial least squares. In : Bagozzi, P. (Ed). Advanced Methods of Marketing Research. Blackwell, 1995, pp. 52-78.

9. Guadagnoli, E, Ward, P. Patient participation in decision-making. Soc. Sci. Med. 1998;47:329-339.

10. Höglund, E. Patienternas erfarenheter : en utgångspunkt för förbättringsarbete. Rapport från studier med Picker-metoden. Stockholm : Spri, 1999. (Spri rapport 488). (In Swedish)

11. Johansson, P-O. Evaluating health risks : an economic approach. Cambridge University Press, 1995.

12. Larsson, G, et al. Refinement of the questionnaire 'Quality of care from the Patients' Perspective' using structural equation modeling. Scandinavian journal of caring sciences 1998;12(2):111-118.

13. Patienterna svarar : patienternas syn på vårdkvaliteten. Stockholm : Spri, 1989. (Spri rapport 271). (In Swedish)

14. Wilde, B., et al. Patienten värderar vården : vägledning till frågeformuläret KUPP, Kvalitet ur patientens perspektiv. Stockholm : Vårdförbundet SHSTF, 1995. (Rapport 45). (In Swedish).

**Appendix**

# Questionnaire

## Examples of questions on "Environment" (page 3 in the 5-page questionnaire)

***Today: The physical standard of the clinic has improved substantially since 1997. However, there is no TV, coffee machine or lockable wardrobe in the clinic waiting room.***

| **How do you grade** | Lowest grade | Highest grade |
|---|---|---|
| - the possibility to find your way to the clinic | | |
| - the homeliness/comfort of the rooms | | |
| - the convenience of the rooms | | |
| - the possibility to deposit your outdoor garment during the visit | | |

***If instead there were:***
- *better sign-posts to the clinic*
- *TV, coffee machine and new magazines in the waiting-room*
- *Lockable wardrobe close to the waiting-room*

| **How would you then grade** | Lowest grade | Highest grade |
|---|---|---|
| - the possibility to find your way to the clinic | | |
| - the homeliness/comfort of the rooms | | |
| - the convenience of the rooms | | |
| - the possibility to deposit your outdoor garment during the visit | | |

***Would you be willing to pay an additional SEK X (10-150) per visit (above the ordinary charge) if***

- there were better sign-posts to the clinic yes no

***Would you be willing to pay an additional SEK X (10-150) per visit (above the ordinary charge) if***

- the waiting-room was equipped with TV, coffee machine and new magazines yes no

***Would you be willing to pay an additional SEK X (10-150) per visit (above the ordinary charge) if***

- lockable wardrobes were installed close to the waiting-room yes no

# EFI
# The Economic Research Institute

**Reports since 1995**
A complete publication list can be found at www.hhs.se/efi

*Published in the language indicated by the title*

*2001*

**Andersson, P.,** Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.
**Björklund, C.,** Work Motivation - Studies of its Determinants and Outcomes.
**Charpentier, C.,** Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämnds-reformen.
**Dahlén, M.,** Marketing on the Web - Empirical Studies of Advertising and Promotion Effectiveness.
**Ekelund, M.,** Competition and Innovation in the Swedish Pharmaceutical Market.
**Ericsson, D.,** Kreativitetsmysteriet – Ledtrådar till arbetslivets kreativisering och skrivandets metafysik.
**Eriksson, R.,** Price Responses to Changes in Costs and Demand.
**Frisell, L.,** Information and Politics.
**Giordani, P.,** Essays in Monetary Economics and Applied Econometrics.
**Gustavsson, P.,** Essays on Trade, Growth and Applied Econometrics.
**Hedlund, A.,** Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.
**Hill, M.,** Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.
**Hvenmark, J.,**Varför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.
**Hägglund, P.B.,** Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.
**Höök, P.,** Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.
**Johansson, C.,** Styrning för samordning.
**Kjellberg, H.,** Organising Distribution - Hakonbolaget and the efforts to rationalise food distribution, 1940-1960.
**Lange, F. och Wahlund, R.,** Category Management – När konsumenten är manager.
**Liljenberg, A.,** Customer-geared competition – A socio-Austrian explanation of Tertius Gaudens.
**Ljunggren, U.,** Nyckeltal i grundskolan i Stockholms stad före och efter stadsdels-nämndsreformen.
**Löf, M.,** On Seasonality and Cointegration.
**Martensen, K.,** Essays on Entry Externalities and Market Segmentation.
**Matros, A.,** Stochastic Stability and Equilibrium Selection in Games.
**Nilsson, A.,** Market Transparency.
**Persson, B.,** Essays on Altruism and Health Care Markets.
**Skoglund, J.,** Essays on Random Effects Models and GARCH.
**Strand, N.,** Empirical Studies of Pricing.
**Thorén, B.,** Stadsdelsnämndsreformen och det ekonomiska styrsystemet - Om budget-avvikelser.

*2000*

**Berg-Suurwee, U.,** Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.
**Bergkvist, L.,** Advertising Effectiveness Measurement: Intermediate Constructs and Measures.
**Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L.,** Koncernchefen och ägarna.
**Bornefalk, A.,** Essays on Social Conflict and Reform.
**Charpentier, C., Samuelson, L.A.,** Effekter av en sjukvårdsreform.
**Edman, J.,** Information Use and Decision Making in Groups.
**Emling, E.,** Svenskt familjeföretagande.
**Ericson, M.,** Strategi, kalkyl, känsla.
**Gunnarsson, J., Wahlund, R., Flink, H.,** Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.
**Hellman, N.,** Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.
**Hyll, M.,** Essays on the Term Structure of Interest Rates.
**Håkansson, P.,** Beyond Private Label – The Strategic View on Distributor Own Brands.
**I huvudet på kunden.** Söderlund, M., (red).
**Karlsson Stider, A.,** Familjen och firman.
**Ljunggren, U.,** Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.
**Ludvigsen, J.,** The International Networking between European Logistical Operators.
**Nittmar, H.,** Produktutveckling i samarbete – Strukturförändring vid införande av nya Informationssystem.
**Robertsson, G.,** International Portfolio Choice and Trading Behavior.
**Schwarz, B., Weinberg, S.,** Serviceproduktion och kostnader – att söka orsaker till kommunala skillnader.
**Stenström, E.,** Konstiga företag.
**Styrning av team och processer – Teoretiska perspektiv och fallstudier.** Bengtsson, L., Lind, J., Samuelson, L.A., (red).
**Sweet, S.,** Industrial Change Towards Environmental Sustainability - The Case of Replacing Chloroflouorocarbons.
**Tamm Hallström, K.,** Kampen för auktoritet – Standardiseringsorganisationer i arbete.

*1999*

**Adler, N.,** Managing Complex Product Development.
**Allgulin, M.,** Supervision and Monetary Incentives.
**Andersson, P.,** Experto Credite: Three Papers on Experienced Decision Makers.
**Ekman, G.,** Från text till batong – Om poliser, busar och svennar.
**Eliasson, A-C.,** Smooth Transitions in Macroeconomic Relationships.
**Flink, H., Gunnarsson, J., Wahlund, R.,** Svenska hushållens sparande och skuldsättning– ett konsumentbeteende-perspektiv.
**Gunnarsson, J.,** Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
**Hamrefors, S.,** Spontaneous Environmental Scanning.
**Helgesson, C-F.,** Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.

**Japanese Production Management in Sunrise or Sunset.** Karlsson, C., (red).
**Jönsson, B., Jönsson, L., Kobelt, G.,** Modelling Disease Progression And the Effect of Treatment in Secondary Progressive MS. Research Report.
**Lindé, J.,** Essays on the Effects of Fiscal and Monetary Policy.
**Ljunggren, U.,** Indikatorer i grundskolan i Stockholms stad före stadsdelsnämndsreformen – en kartläggning.
**Ljunggren, U.,** En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
**Lundbergh, S.,** Modelling Economic High-Frequency Time Series.
**Mägi, A.,** Store Loyalty? An Empirical Study of Grocery Shopping.
**Mölleryd, B.G.,** Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
**Nilsson, K.,** Ledtider för ledningsinformation.
**Osynlig Företagsledning.** Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red).
**Rognes, J.,** Telecommuting – Organisational Impact of Home Based – Telecommuting.
**Sandström, M.,** Evaluating the Benefits and Effectiveness of Public Policy.
**Skalin, J.,** Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
**Spagnolo, G.,** Essays on Managerial Incentives and Product-Market Competition.
**Strauss, T.,** Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
**Svedberg Nilsson, K.,** Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
**Söderström, U.,** Monetary Policy under Uncertainty.
**Werr, A.,** The Language of Change The Roles of Methods in the Work of Management Consultants.
**Wijkström, F.,** Svenskt organisationsliv – Framväxten av en ideell sektor.

*1998*

**Andersson, M.,** On Testing and Forecasting in Fractionally Integrated Time Series Models.
**Berg-Suurwee, U.,** Styrning av kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.
**Berg-Suurwee, U.,** Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.
**Bergström, F.,** Essays on the Political Economy of Industrial Policy.
**Bild, M.,** Valuation of Takeovers.
**Charpentier, C., Samuelson, L.A.,** Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.
**Eriksson-Skoog, G.,** The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.
**Gredenhoff, M.,** Bootstrap Inference in Time Series Econometrics.
**Ioannidis, D.,** I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.
**Johansson, S.,** Savings Investment, and Economic Reforms in Developing Countries.
**Levin, J.,** Essays in Company Valuation.
**Ljunggren, U.,** Styrning av grundskolan i Stockholms stad innan stadsdelsnämndsreformen.
**Mattsson, S.,** Från stat till marknad – effekter på nätverksrelationer vid en bolagiseringsreform.
**Nyberg, A.,** Innovation in Distribution Channels – An Evolutionary Approach.
**Olsson, P.,** Studies in Company Valuation.
**Reneby, J.,** Pricing Corporate Securities.

**Roszbach, K.,** Essays on Banking Credit and Interest Rates.
**Runsten, M.,** The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.
**Segendorff, B.,** Essays on Bargaining and Delegation.
**Sjöberg, L., Bagozzi, R., Ingvar, D.H.,** Will and Economic Behavior.
**Sjögren, A.,** Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.
**Studier i kostnadsintäktsanalys.** Jennergren, P., (red)
**Söderholm, J.,** Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.
**Thorburn, K.,** Cash Auction Bankruptcy and Corporate Restructuring
**Wijkström, F.,** Different Faces of Civil Society.
**Zethraeus, N.,** Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.

*1997*

**Alexius, A.,** Essays on Exchange Rates, Prices and Interest Rates.
**Andersson, B.,** Essays on the Swedish Electricity Market.
**Berggren, N.,** Essays in Constitutional Economics.
**Ericsson, J.,** Credit Risk in Corporate Securities and Derivatives. Valuation and Optimal Capital Structure Choice.
**Charpentier, C.,** Budgeteringens roller, aktörer och effekter. En studie av budget-processerna i en offentlig organisation.
**De Geer, H., Silfverberg, G.,** Citizens'Trust and Authorities' Choices, A Report from The Fourth International Conference on Ethics in the Public Service, Stockholm June 15-18, 1994.
**Friberg, R.,** Prices, Profits and Exchange Rates.
**Från optionsprissättning till konkurslagstiftning.** Bergström, C., Björk, T., (red)
**Hagerud, G.E.,** A New Non-Linear GARCH Model.
**Haksar, A.,** Environmental Effects of Economywide Policies: Case Studies of Costa Rica and Sri Lanka.
**He, C.,** Statistical Properties of Garch Processes.
**Holmgren, M.,** Datorbaserat kontrollrum inom processindustrin; erfarenheter i ett tidsperspektiv.
**Jennergren, P.,** Tutorial on the McKinsey Model for Valuation of Companies.
**Lagerlöf, J.,** Essays on Political Economy, Information, and Welfare.
**Lange, F., Wahlund, R.,** Planerade och oplanerade köp - Konsumenternas planering och köp av dagligvaror.
**Löthgren, M.,** Essays on Efficiency and Productivity; Contributions on Bootstrap, DEA and Stochastic Frontier Models.
**Nilsson, H.E.,** Using Conceptual Data Modelling in Management Accounting: A Case Study.
**Sjöberg, L., Ramsberg, J.,** En analys av en samhällsekonomisk bedömning av ändrade säkerhetsföreskrifter rörande heta arbeten.
**Säfvenblad, P.,** Price Formation in Multi-Asset Securities Markets.
**Sällström, S.,** On the Dynamics of Price Quality.
**Södergren, B.,** På väg mot en horisontell organisation? Erfarenheter från näringslivet av decentralisering och därefter.
**Tambour, M.,** Essays on Performance Measurement in Health Care.
**Thorén, B., Berg-Surwee, U.,** Områdesarbete i Östra Hökarängen - ett försök att studera effekter av decentralisering.
**Zhang Gang.,** Chinese Rural Enterprises Between Plan and Market.
**Åhlström, P.,** Sequences in the Profess of Adopting Lean Production.

**Åkesson, G.,** Företagsledning i strategiskt vakuum. Om aktörer och förändrings-processer.
**Åsbrink, S.,** Nonlinearities and Regime Shifts in Financial Time Series.

*1996*

**Advancing your Business. People and Information Systems in Concert.**
Lundeberg, M., Sundgren, B (red).
**Att föra verksamheten framåt. Människor och informationssystem i samverkan.**
red. Lundeberg, M., Sundgren, B.
**Andersson, P.,** Concurrence, Transition and Evolution - Perspectives of Industrial Marketing Change Processes.
**Andersson, P.,** The Emergence and Change of Pharmacia Biotech 1959-1995. The Power of the Slow Flow and the Drama of Great Events.
**Asplund, M.,** Essays in Industrial Economics.
**Delmar, F.,** Entrepreneurial Behavior & Business Performance.
**Edlund, L.,** The Marriage Market: How Do You Compare?
**Gunnarsson, J.,** Three Studies of Financial Behavior. Research Report.
**Hedborg, A.,** Studies of Framing, Judgment and Choice.
**Holgersson, C., Höök, P.,** Ledarutveckling för kvinnor - Uppföljning av en satsning på Volvo - Research Report.
**Holmberg, C.,** Stores and Consumers - Two Perspectives on Food Purchasing.
**Håkansson, P., Wahlund, R.,** Varumärken. Från teori till praktik.
**Karlsson, A.,** The Family Business as an Heirloom. Research Report.
**Linghag, S.,** Man är handelsstudent. Research Report.
**Molin, J.,** Essays on Corporate Finance and Governance.
**Mägi, A.,** The French Food Retailing Industry - A Descriptive Study.
**Mölleryd, B.,** Så byggdes en världsindustri - Entreprenörskapets betydelse för svensk mobiltelefoni. Research Report.
**Nielsen, S.,** Omkostningskalkulation for avancerede produktions-omgivelser - en sammenligning af stokastiske og deterministiske omkost-ningskalkulationsmodeller.
**Normark, P., Danielsson, L., Larsson, A., Lundblad, P.,** Kooperativa nyckeltal. Research Report.
**Sandin, R.,** Heterogeneity in Oligopoly: Theories and Tests.
**Sandén-Håkansson, U.,** Från kampanjmål till mediemix - en studie av samarbete mellan annonsörer, reklambyråer och mediebyråer. Research Report.
**Stein, J., Söderlund, M.,** Framgång i arbetet, strategier för att utföra arbetsuppgifterna, arbetsuppgifternas karaktär och utbildningskvalitet. En empirisk studie av civil-ekonomer. Research Report.
**Strömberg, P., Thorburn, K.,** An Empirical Investigation of Swedish Corporations in Liquidation Bankruptcy. Research Report.
**Söderlund, M.,** Och ge oss den nöjda kunden. En studie av kundtillfredsställelse och dess orsaker och effekter. Research Report.
**Thodenius, B.,** Användningen av ledningsinformationssystem i Sverige: Lägesbild 1995. Research Report.
**Ulfsdotter, U.,** Internationalisering för expansion eller hemmamarknadsförsvar? De nordiska marknaderna för fruktyoghurt 1982-1994.
**Westelius, A.,** A Study of Patterns of Communication in Management Accounting and Control Projects.
**Wijkström, F.,** Den svenska ideella sektorn och pengarna. Research Report.
**Örtendahl, M.,** Health and Time - A Problem of Discounting.