NOTAS ECONÓMICAS 1

REVISTA DA FACULDADE DE ECONOMIA DA UNIVERSIDADE DE COIMBRA

NÚMERO 1 / ABRIL'93 / PREÇO 1 500$00 / ISSN 0872-4733

The use of the multiattribute utility theory
to measure the stability of patients' quality judgments

Pedro Lopes Ferreira Professor da Faculdade de Economia da Universidade de Coimbra.

resumo

Neste artigo, o autor pretende descrever uma investigação levada a cabo para saber se as opiniões dos doentes sobre a qualidade dos cuidados hospitalares dependem do momento em que são emitidas. Para conhecer essas opiniões, mediu-se a satisfação, utilizando escalas do tipo Likert, e as expectativas ao longo do tempo, através da aplicação da teoria multi-atributo de utilidade.

Para este trabalho de investigação, seleccionaram-se, em nove hospitais, doentes diagnosticados com enfarte do miocárdio. Com o auxílio de questionários, coligiram-se dados referentes às respostas de 164 doentes, um mês e cinco meses após a alta.

Os resultados mostraram que, pelo menos nos casos de enfarte do miocárdio, a satisfação e as expectativas dos doentes não são necessariamente estáveis no tempo. Conclui-se que o momento da fase de recuperação em que as avaliações são pedidas é um elemento fundamental.

resumé / abstract

Dans cet article, l'auteur se propose de décrire une recherche qui été menée dans l'objectif de déterminer si les évaluations faites par les malades sur la qualité des soins hospitaliers dépendent du moment où elles sont émises. Pour connaître ces évaluations, on a mesuré le degré de satisfaction à l'aide d'échelles de Likert, et les expectatives au long du temps à l'aide de l'application de la théorie des multi-attributs de l'utilité.

Pour ce travail de recherche, on a sélectionné dans neuf hôpitaux, des malades souffrant d'infarctus du myocarde. A l'aide de questionnaires, on a recueilli les données référentes aux réponses de cent soixante-quatre malades, un mois, puis cinq mois après leur autorisation de sortie de l'hôpital.

Les résultats démontrent que, tout au moins dans le cas d'infarctus du myocarde, la satisfaction et les expectatives des malades ne sont pas nécéssairement stables dans le temps. On en conclut que le moment de la phase de rétablissement où ces évaluations sont demandées est un élément fondamental.

The purpose of this paper is to describe a research carried out to determine whether patients' judgments concerning quality of hospital care depend on when their judgments are assessed. To obtain patients' judgments about quality of care the author measured their satisfaction, by using traditional Likert-type scales, as well as their expectations over time, using the multiattribute utility approach.

The current research had a within-subject design and was diagnosis-specific. People diagnosed with myocardial infarction were selected from nine hospitals. Through self-administered questionnaires, data were collected from one hundred sixty-four patients at two points of time: one month and five months after discharge.

Results showed that, at least for myocardial infarction, patients' satisfaction and expectations are not necessarily stable over time. The time at which the assessment is administered during the patient's recovery process is crucial.

1 - Introduction

Some authors have presented techniques for incorporating a person's values into the decision-making process (Beach *et al.*, 1976; Edwards, 1977; Hammond *et al.*, 1980) particularly in clinical areas (McNeil *et al.*, 1978; 1981; Moles, 1982). However, little research has been done regarding the stability of patients' values and preferences. Christensen-Szalanski (1984) showed that patients' preferences vary over time and also that during certain periods of time patients' values may not be representative of their long-term preferences. His study included pregnant women and he analyzed their attitudes toward avoiding pain and anesthesia one month pre- and one month post-partum.

Ware *et al.* (1983) developed a self-administered survey instrument designed for use in general population studies. This questionnaire contained items which addressed variables related to characteristics of doctors and medical services (e.g., technical and interpersonal skills of providers, waiting time for appointments, office waits, emergency care, costs of care, insurance coverage, and availability of hospitals). These researchers administered the questionnaires twice, approximately two years apart, and found that satisfaction was relatively stable over time.

The contradiction showed by these two studies reveals one of the reasons we decided to explore the intertemporal stability in a disease that has a significant impact on patients' lives. We defend that, at least until further research, we cannot compare satisfaction studies for the general population with those performed on specific cohorts of patients. On the other hand, Ware's findings cannot be extrapolated outside his study because the questionnaires used were not identical on both administrations. In fact, regarding the different aspects of quality of care, Nelson *et al.* (1989) identified the need for assessing patients' perceptions about quality as an issue that should be further researched.

In this study we measured patients' expectations regarding the quality of hospital care in general as well as patient satisfaction after treatment of a particular medical problem. Expectations for each area of care were captured by the weights and the utility functions patients provided at both points in time. The weights gave us the relative importance of each of the attributes of care. The utility functions gave us the psychological value of different levels of each attribute.

Using patients' scores evaluating the different areas of care we also obtained patients' evaluations of a specific visit to a specific hospital as a measure of how their expectations were met. A 5-point Likert-type scale from *excellent* to *poor* was used as the way to measure patient satisfaction.

2 - Importance of this Study

This study was important because it analyzed how patients' expectations changed over time. If we want to incorporate patients' perceptions and judgments in a model to measure quality of care, we must be aware of the intertemporal stability of these perceptions and judgments. This is also a fundamental question concerning all utility models: does the assessment period matter? Are patients' scores, weights and utilities stable over time?

This research aimed at improving the interpretation of measures of patient satisfaction by incorporating the time dimension. The difference between health measures at two different points in the patient recovery process can represent health benefits. This difference can also provide insights into improving the health care system.

The implications of measuring satisfaction and expectations over time should also be highlighted as a contribution of this research. Concluding that satisfaction scores are stable over time does not immediately lead us to any conclusion about the stability of patients' judgments. If patients' expectations did not change over time, then we may presume that patients were in fact stable, because both satisfaction scores and expectations were stable. On the other hand, if the expectations were unstable, we may have a situation were a patient rated an area of care identically at both points in time, but with completely different value judgments.

By using only satisfaction scores, it is impossible to detect some situations were patients' expectations changed over time. A similar problem occurs when satisfaction scores are not stable. If patient's expectations are also not stable, we may be more comfortable with the idea of the intertemporal instability of these patients. But, if the expectations are stable over time and the satisfaction scores are not, then a change in satisfaction may emerge from a mere change of the perception of the situation.

The results of this study are also important in terms of clinical practice. This is because patient compliance with treatment has been shown strongly dependent on how satisfied the patient is with the quality of the physician's work.

In addition, the better the hospital meets patients' needs and wants, the higher is the probability for patients to comply with the treatment and to return and recommend the same hospital to others (Nelson *et al.*, 1989). Understanding the process of the stability over time of patients' judgments is the first step towards a full understanding of patients' needs and wants.

3 - Background Research

The background for this study's design was mainly the available literature on how quality of care has been defined, how patient satisfaction has been defined and assessed, and how utility theory has been used in health care to assess individuals' values about quality. Two other important ingredients were the time dimension and patients' perceptions of their own health status.

Quality of care is a term that is rarely defined by those using it. Yet everyone claims that he understands what it means (Kincaid, 1981). One reason why quality of care is difficult to define and measure is because we may have numerous viewpoints when assessing this concept. Whenever we talk about quality, we should specify what we mean by quality: which aspect of quality, quality for whom and for what purpose, and quality by whom.

We used Donabedian's definition of quality of care as "that kind of care which is expected to maximize an inclusive measure of patient welfare, after one has taken account of the balance of expected gains or losses that attend the process of care in all its parts" (Donabedian, 1980: 5-6).

Since research in this area began, it has been difficult to distinguish the definition of the concept of quality of care from its operationalization. All researchers in this field agree that to develop a usable definition of quality of care it is necessary to enumerate the elements which belong to it. We need to operationalize the definition of quality of care. Donabedian summarized this need when he wrote that doing any measurement without an accurate definition of what quality of care means, is "to court disaster".

It emerges from the literature that there are five axes used by patients to measure the quality of care, and eventually patient satisfaction (Donabedian, 1980; Hulka *et al.*, 1971, 1975; Vuori, 1982; Ware and Snyder, 1975; Ware *et al.*, 1983). They are: (1) the technical component of care or "curing" function; (2) the interpersonal aspect of care or "caring" function; (3) the accessibility; (4) the availability; and (5) the continuity of care.

3.1 Framework to measure patient satisfaction

Patient evaluation of the quality of care and patient satisfaction can be rooted in the decision science framework. We applied the multiattribute utility technology in our research and focused our attention on only one family of models: the multiattribute utility (MAU) models (Fishburn, 1977; Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). They constitute a generic analytic structure a decision maker or analyst may use to model a problem. They also have been "a widely accepted and frequently applied tool for assisting decision makers in making choices among complex alternatives that vary on multiple conflicting objectives" (Borcherding *et al.*, 1989). In the multiattribute utility theory (MAUT), the evaluation task is broken down into attributes, and single-attribute evaluations are constructed. The tradeoffs among attributes are quantified as importance weights or other scaling factors.

Because the multiattribute utility approach can deal with the utility of any quantifiable good, it makes sense to use this theory in health care. In fact, in health care, the family of MAU models are the most popular way, especially in the health status literature (Patrick *et al.*, 1973; Torrance, 1986; Torrance *et al.*, 1982), to capture patients' judgments, to measure the value of life (Pliskin *et al.*, 1980), or to evaluate treatments (Krischer, 1976).

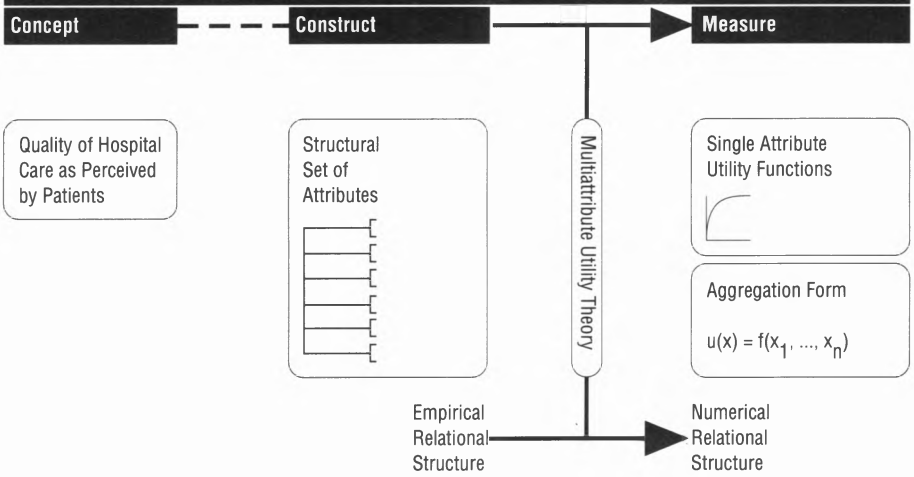
Other studies have been using the same type of models. For instance, the majority of indices in health care use summating techniques (additive MAU models) to obtain final scores (Gustafson *et al.*, 1990; Hughes *et al.*, 1988; Nelson *et al.*, 1989). Multiattribute utility indices have also been shown to be highly reliable, valid and useful in assessing severity (Gustafson and Johnson, 1988). As these models usually are generally evaluating feelings of favorableness or unfavorableness towards an object, this type of model is also justified by psychological theories (Fishbein and Ajzen, 1975; Linder-Pelz, 1982).

Usually, four tasks are involved in a multiattribute utility measurement. One of the first steps whenever we want to evaluate a problem involving multiattributes is to solicit the set of attributes and to structure the problem in a value tree, a structured list of attributes. We also need to operationalize these attributes. The next step is the elicitation and

construction of single attribute utility functions which preserve the revealed preferences of the decision maker. The final steps address the problem of eliciting weights of attributes and aggregating values or utilities across attributes.

Figure 1 presents our framework to measure patient satisfaction. Initially, we want to measure the concept of quality of hospital care as perceived by patients. As it constitutes an abstract and complex concept, the best thing we can do is to define (build) a valid and reliable construct (set of structural dimensions) for our specific concept. Our next concern is to obtain meaningful measures. So, using the multiattribute utility theory, we are able to come up with a set of single attribute utility functions and an appropriate mathematical form to obtain an overall utility.

Figure 1 - Framework to Measure Patient Satisfaction



3.2 Quality dimensions

The study attributes for this research correspond to the sequence of events that patients experience as they enter the hospital, receive treatment and are discharged. They constitute a construct which is intended to represent the concept of quality of care as perceived by patients (see Table 1).

These six quality scales were based on items from the Patient Judgments of Hospital Quality (PJHQ) questionnaire (Meterko *et al.*, 1990). The authors of this instrument assume that, when patients rate hospital care, they distinguish among kinds of staff and services. They also defended the clear independence of these dimensions, suggesting that patients' evaluations should be assessed and interpreted separately.

Table 1 - Semantics of Quality Scales

Admitting	Efficiency of the admiting procedure. Courtesy of admitting staff to patients' individual needs' comfort and feelings.
Nursing and daily care	Nurses' caring and curing. Courtesy and respect, friendliness and kindness Skills of nurses.
Medical care	Doctors' caring and curing. Courtesy and respect, friendliness and kindeness. Their ability to diagnose problems, thoroughness of examinations, skill in treating patients' conditions, and scientific knowledge.
Information	Information given by nurses about daily procedures and by doctores about illness and treatment.
Hospital rooms	Amount of peace, quiet and comfort.
Discharge	Discharge procedure efficiency and coordination of care after discharge.

Each scale's internal consistency and homogeneity have been studied. As a corollary of these studies, researchers have found reliability estimates (Cronbach's alpha) good or excellent. All of them exceeded the minimum 0.70 criterion Nunnally (1978) recommended for hypothesized constructs. They ranged from 0.87 to 0.95 with four exceeding 0.90. Homogeneity estimates were acceptable and ranged from 0.47 to 0.69. Construct validity was also analyzed. The correlations among the six hospital process scales were found to be moderate to strong, positive and statistically significant. Also the correlations between the six hospital process scales and other variables to which they should be related to (e.g., overall quality, recommendations and intentions, and overall health outcomes) were moderate to strong, ranging from 0.39 to 0.75 (median r=0.51). Lastly, a total of 63% of the variance in overall quality of care and 47% of the variance in recommendations and intentions were accounted for by the six process scales.

3.3 Unidimensional utility functions

To express how desirable an option is with respect to a particular attribute, we usually use the concept of utility. A unidimensional utility function u_i corresponding to the i^{th} attribute is defined by $u_i(x_i) = u(x_i, \overline{x_i^0})$. This means that the other attributes are at some fixed level.

In this study, by assessing utility functions, we implicitly assessed the patients' preferences for different levels of each attribute of quality of hospital care. These preferences were mapped into a 0-100 scale, where 0 corresponded to the worst possible quality, and 100 to the best possible one. Intermediate levels were assigned by the patient according to his relative preference.

To assess utility functions, four scaling techniques have been commonly used in health-related measurement: (1) Category Scaling, (2) Magnitude Scaling, (3) Standard Gamble, and (4) Time Trade-Off technique (Froberg and Kane, 1989)

Category Scaling requires that patients place marks on a line ranging from 0 (worst quality, death) to 100 (best quality, full healthy life). Magnitude Estimation is similar to category scaling except that just one reference state is required, and proportional quality states are made. Standard Gamble involves patients in choices between certain and uncertain quality outcomes. Lastly, the Time Trade-Off method, more frequently utilized in health status measurements, presents patients with choices concerning the duration of different states.

There is no clear evidence that one method is better than other. In fact, the standard gamble method has its foundations in utility theory (von Neumann & Morgenstern, 1944). It is based on axioms and incorporates a conceptual framework for decision making under uncertainty. However, this is also one of its drawbacks, especially after the axioms of expected utility theory had been called into question in empirical testings. Also, some patients find this technique difficult to understand and others resist gambling with health. In health status assessment, Torrance used this technique in the context of a multiattribute utility function. He used a probability wheel to help patients understand the procedure (Torrance, 1986). This tool is a disk with two movable different-colored sections which are adjusted to represent the probabilities p and 1-p of the two gamble alternatives.

The time trade-off method (Drummond *et al.*, 1987; Torrance *et al.*, 1982) lacks the theoretical properties of the expected utility approach (no risk involved) and requires several iterations difficult to be implemented in a self-administered questionnaire. However, it can be used with success for a chronic health state considered better than death, chronic health state considered worse than death, or temporary health state. In this research all these advantages are irrelevant, compared to the inherent difficulty of utility assessment using a mailed questionnaire.

The category scaling method, using a line with two clearly defined anchors, is relatively easy for patients to understand. Its major advantage is its practicality. Strictly speaking, however, this technique leads to an ordinal scale of utilities. Torrance (1986), however, argue that the standard gamble, time trade-offs and category ratings all require subjects to respond in terms of an interval scale.

To obtain a ratio scale, some authors defend the use of the magnitude estimation (Froberg and Kane, 1989), procedure based on S. S. Stevens' work. Kaplan and Ernst (1983) for instance, stated that "the mean category ratings are linearly related to the logarithms of the arithmetic or geometric mean magnitude estimation judgments." The discussion around the endpoint zero is still open. Some authors address this issue, for health status, defining zero as the absence of dysfunction and discomfort; others point out that this procedure turns a health status index into an "illness index".

It seems to us, however, that cognitively this final technique is slightly more difficult to implement than the category scaling. We then decided to use the category scaling technique to assess utility functions for quality of hospital care. It appears to provide scale values that are as valid as the ones obtained by magnitude estimation or time trade-off methods.

3.4 Importance Weights and Aggregation Form

To obtain the aggregation of the attributes, we usually need to know the relative importance weights of each of them. The major weighting methods for riskless outcomes existing in multiattribute utility measurement are: the ranking method, the ratio method, the direct rating method, and the swing weighting method. All of these methods use numerical estimation procedures.

We only mention these four methods because they were the only candidates we considered for this study. For instance, the tradeoff method (Keeney and Raiffa, 1976) was not accepted because it is difficult to implement in self-administered questionnaires and it is time consuming. The pricing out method was also excluded because of the difficulty of assigning dollar values to satisfaction.

In the ranking method (von Winterfeldt and Edwards, 1986), the subjects are asked to rank the different attributes in the order of their importance. In the ratio method, the weights are elicited hierarchically. The subject begins at the bottom of the value tree and gives a score of 10 to the least important attribute and assigns scores to the other attributes as multiples of 10.

In the direct rating method, subjects are asked to directly assign numbers to attributes in order to quantify their relative importance. A modification of this method consists of asking subjects to distribute 100 points over the attributes. This may be considered a variation of the ratio estimation procedure (von Winterfeldt and Edwards, 1986).

The swing weighting method (von Winterfeldt and Edwards, 1986) is a non-hierarchical one. As a stimulus the subjects receive all the attributes of the value tree on their lowest level. The subjects are then asked which "swing" from the lowest to the best provides a better improvement on the overall value or utility. They continue to be asked for the swing allowing a second better improvement, and so on. At the end we have a rank order of the weights. At this point, they are asked to assign 100 to the largest swing and express the magnitude of the other swings in terms of percentage of the largest one. As in the tradeoff method, the swing weighting method is difficult to implement in self-administered questionnaires, especially in an already long questionnaire.

The final information to be presented in this section is the mathematical aggregation form. The most common ways to aggregate dimensional scales are the weighted additive model and the multiplicative model. The common feature of these forms of decomposition is that they are simple polynomials, and decomposition theorems, under certain conditions, guarantee that these decompositions are possible. An introduction to these models is presented in Keeney and Raiffa (1976) and a survey of multiattribute utility theory and applications are included in Farquhar (1977).

A substantial portion of multiattribute utility theory deals with weighted additive models:

$$u(\mathbf{x}) = \sum_{i=1}^{n} w_i u_i(x_i)$$

where $\mathbf{x}$ is the evaluation object; x_i is its measurement on attribute i; u_i is the single-attribute utility function; w_i is the weight of attribute i; u is the overall value of $\mathbf{x}$; and n is the number of attributes. We assume here that the whole is equal to the sum of the parts, and that the contribution of each scale attribute is independent of the values of the remaining attributes. This type of model has been studied by several authors, especially in nonrisky decision problems.

The other common model is the multiplicative model:

$$1+ku(\mathbf{x}) = \prod_{i=1}^{n} \{1 + kw_i u_i(x_i)\}$$

where k defines the interaction terms, and the other components were previously defined.

As the number of interacting terms increases, the power of k also increases. The value of k must lie between -1 and infinity. As |k| becomes larger, the overall interaction among all attributes becomes more heavily weighted.

The idea of using weights is to express the importance of each attribute relative to all others. Keeney and Raiffa (1976) discussed how difficult it is to associate weights with the concept of importance. They considered weights as mere rescaling constants used to match the units of one single-attribute function with the units of another. Weights are usually sensitive to the range of these functions' scales; weights increase (decrease) when the range of the single-attribute function increase (decrease). Von Winterfeldt and Edwards (1986) also suggested that we should use swing weighting as an alternative solution to the range problem. Using an appropriate anchored single-attribute elicitation technique, the swing weighting is "virtually indistinguishable" from the other difference measurement techniques.

However, the weights capture the essence of value judgments. Values are reflected in weights, and values change over time. So, weights should be re-elicited in situations in which a program is periodically evaluated.

4 - Research Hypotheses

The general research question was: *Do patients' judgments of quality depend on when they are assessed?* But, as mentioned before, this research used two types of quality judgments: (1) satisfaction scores regarding the performance of the hospital in various dimensions of care and measured by a 5-point scale; and (2) patients' expectations measured by the importance weights they assign to the different areas of care and by the component utility functions for each of these areas. Therefore, based on the conceptual framework presented above and on the review of the literature on quality of care and on patient satisfaction, we formulated the questions presented on Table 2.

Table 2 - Research Questions

Q1	Stability of scores - Are the scores stable over time?
Q2	Stability of Weights - Are the weights over time?
Q3	Stability of utilities - Are the utility values stable over time?

5 - Methods

This research was an observational study with a simple repeated measures design. Nine community hospitals located in four regions of the United States participated in the study: Florida (three hospitals), Georgia (three hospitals), South Carolina (one hospital), and Tennessee (two hospitals). All of these hospitals were managed by a major national proprietary hospital chain and constituted a homogeneous sample. In fact, all nine were community hospitals, a type of hospital where little work has been done, if compared with major medical centers.

Patients discharged alive from these nine hospitals with myocardial infarction (MI) as their primary discharge (principal) diagnosis, were asked to answer two questionnaires at two different points in time. We decided to use the interval 1-2 months post-discharge for the first time period, and the interval 4-5 months post-discharge for the second one. Early ratings (immediately after discharge) are said to reflect patients' gratitude for being alive and were therefore not used in this study.

The myocardial infarction diagnosis was selected after defining a set of three criteria. These criteria included the existence of previous studies, the high volume and the wide variation in satisfaction and outcomes. MI was also chosen because the process of care related to it is very standard (Wennberg, 1987). Moreover, the discharge information (one of the dimensions studied) is very important in MI patients.

The empirical study for this research began in April 1990 when the first hospitals were invited to participate. We had three forms of collecting data: (1) self-administered questionnaires for perceived health status, quality utility data, disease specific symptoms, and satisfaction with care; (2) short telephone interviews to remind non-respondents and/or to obtain the data for a small subset of questions; and (3) medical records and clinical data from hospital data centers, especially on demographic and descriptive characteristics.

Completion of each questionnaire took approximately 20 minutes. Patients who had not returned completed questionnaires were sent a post card two weeks after the original mailing. This reminded them to complete the questionnaire if they wished to do so.

The telephone interviews were used mainly to remind again a patient that he had not answered the corresponding

questionnaire. However, whenever possible, an actual interview of less than ten minutes took place. In these interviews, the main questions asked were those related to how the patients' perceived health status, how they evaluated, in general, the quality of health care provided to them by the hospitals (patients' expectations) and how they evaluated their particular stay in the hospital (satisfaction scores). Three attempts were made to contact non-respondent patients.

Patients' demographic and descriptive characteristics were obtained from the hospitals' medical records. These variables included hospital and patient numbers, patient's name, address, and telephone number as well as gender, marital status, date of birth, primary discharge (principal) diagnosis, DRG, and length of stay.

From the eleven hospitals invited, nine agreed to participate in this study. Hospital leaders from the nine selected hospitals were then involved from the beginning of this study. Hospital administrators were contacted, as were some physicians and the hospital "quality coach."

An independent research firm, with extensive experience in providing patient quality measurement feedback systems, helped in printing the questionnaires, in selecting the sample and in reviewing and entering the data. The questionnaires were mailed out in hospital envelopes and the return envelopes were addressed to that research firm.

Risks to the subjects were minimal. There were no requirements of them other than the time it took to complete the questionnaires. Psychological risks were also minimal since each patient was able to express his opinion. Any subject who did not wish to participate or continue with the study, for whatever reason, was free to withdraw at any time. This did not affect his hospital care in any way. Individual subjects were not likely to gain personally from participating in this study unless participation fill some personal needs, such as a desire to verbalize thoughts related to the hospital care or feelings of satisfaction related to participating in a study.

In the nine hospitals, all eligible cardiac cases were selected. The final selection criteria for including a patient's case in our study was that he had been discharged with one or more of the diagnoses codes ICD-9-CM 410.x and 411.x, and had stayed in the hospital for at least five days (otherwise they are not clinically considered "real" Mis). Excluded were cases of death during the hospitalization or during the first 1-2 months after discharge. Patients less than one year of age or more than 80, with any diagnosis of mental disorder, alcohol or drug dependence, or brain disorder were also excluded from this study, as well as patients who had been discharged against medical advice.

In order to have a sample not contaminated by other problems, we excluded other diagnoses usually related to MI as for instance, old myocardial infarction (ICD-9-CM code 412). This so-called silent MI might have confounded the sample, because it corresponds to patients who have been diagnosed on electrocardiogram (ECG) as having had a myocardial infarction in the past, but who currently present no symptoms. Also, other specific forms of chronic ischemic heart disease (ICD-9-CM code 414.8) were excluded. They correspond to people who previously had an acute myocardial infarction and are admitted with symptoms. They may or may not have an MI.

Names of potential subjects were obtained from the hospital data centers. From our starting sample of size 164 we obtained a 80.5% response rate. A total of 111 (67.7%) patients completed the questionnaires, 21 (12.8%) were telephone interviewed during the first data collection, and 32 were non-respondents. Five months later, we sent a second questionnaire to the 132 respondents on the first data collection. 80 (60.6%) of these patients completed and mailed the second questionnaire, 32 (24.2%) were telephone interviewed and 20 (15.1%) were declared non-respondents. It was not possible to ascertain whether patients who did not respond were still living. The overall response rate was 68.3% (112/164).

The majority of the 132 respondents were men and were married. Almost half of them were over 65 years old. The average age among the respondents was 62.2 (S. D. = 10.7) years. The study patients stayed an average of 9.5 (S. D. = 4.8) days in the hospital and 45% of them had previously been hospitalized for MI. ICD-9-CM codes 410.1, 410.4, 410.7, and 411.1 were the most common diagnosis. Finally, 48% of the patients were Medicare recipients, and 27% and 14% respectively belonged to major commercial insurers and to Blue Cross.

These background variables were tested for differences between respondents and non-respondents. None of the variables differed significantly between both groups of patients. For the nominal measures (gender, marital status, and payor status) we used the chi-square test for rx2 tables. Differences between means of interval measures (age, and length of stay) were tested by the t-test.

6 - Results and Discussion

We studied the percent of patients who moved from one particular satisfaction score at time 1 to another satisfaction score at time 2. We also performed matched pair Wilcoxon tests to study the difference of the distributions of the scores at both times, as well as matched pair t-tests to study how significantly from zero the mean differences of satisfaction score were at both times.

Answering the first research question, we concluded that, in spite of having a majority of stable patients (from 54.4% to 81.4% with an average of 66.0%), we obtained a substantial number of patients unstable and generally lesser satisfied five months after discharge that they were at discharge. The attributes related to time spent on paperwork at both admission and discharge showed a statistically significant change on the level of dissatisfaction from time 1 to time 2. Except for the attributes related to skills and information provided by doctors, we found a decrease in satisfaction five months post-discharge.

To better understand patients' perceptions of the relative importance of the dimensions of care, for each attribute we started by tabulating the rank orders. We performed the Wilcoxon test to determine whether the time 1 and time 2 distributions for each dimension of care differed. None of the distributions significantly differed over time.

Then, for each patient, we studied the changes over time of the rank orders of the dimensions of care. The scale used allow us an ordinal and an interval scale measurement within subjects. Therefore, to analyze the changes of rank orders over time, we computed the average sum of squared distances between the scores at both time periods. Answering the second research question, we concluded that a large majority of patients did not show stability when the rank order of dimensions of care was assessed. We also used the Kendall rank-order correlation coefficient t and we obtained approximately the same results.

In terms of functional forms of the utility functions, we obtained interesting findings. The majority (60% to 89%) of the patients gave, at both time periods, utility functions with functional forms not significantly different from the diagonal. The only exception is the attribute 'restfulness of the room' where 97.8% (time 1) and 95.5% (time 2) of the patients provided us with non-monotonic functions. These patients did not want to feel lonely or ignored; a room where it is always easy to rest is as bad as a room where it is never easy to rest.

Among the exceptions, we found non-monotonic functions for the attributes 'time on paperwork/admission' and 'time of the discharge planning'. They considered that a minimum optimal time is always required on admission and to receive instructions at discharge. If that minimum time is not achieved then patients consider that their expectations are not met. If that optimal point is largely exceeded, the same patients consider that they are wasting their time without corresponding benefits. Again within the exceptions, all the remaining attributes (except the attribute 'skill of nurses') had a convex curve which reveals their importance to the patients.

In summary, it seemed that the utility functions could be considered as stable over time. However, when the stability was analyzed attribute by attribute we obtained very high percentages of unstable cases. Except for the attribute 'restfulness of room', where the majority of patients were very stable giving non-monotonic functions, we always obtained percentages of instability ranging from 16.3% to 42.6%. These percentages are large enough to reject the hypothesis of stability of the functional form of the utility functions. More that 25% of the utilities assessed were instable. These findings by themselves evidenced the presence of instability on the utility functions, when these are elicited regarding the quality of hospital care.

Finally, we performed the multivariate analysis Hotelling's T^2 on the actual numbers given by the patients. Only the attributes 'courtesy of admitting staff' and 'paperwork on discharge' were significant in the multivariate analysis.

7 - Conclusion

This research showed that, at least for myocardial infarction patients, patients' expectations (weights and utilities) are not necessarily stable over time. The major implication of this is that every time we decide to use a utility model to measure the quality of the hospital care, we have to study the appropriateness of this model at that point in time. The time of assessment during the patient's recovery process is crucial.

Practically speaking, these findings create some uncertainty regarding the best time to assess patients' quality judgments as well as the type of the initiatives to develop to meet patients' wants and needs. Hospitals must take the intertemporal stability of their patients' judgments into account. To fully address this issue we need further research.

We need to understand what kind of patients are unstable over time and what implications that instability has on the effectiveness of efforts toward process improvement and decision making. We need to replicate this study to other diagnoses to be able to focus our future research on the patients who show less stability. The hospitals also need to define for how long after the discharge they want their patients to be satisfied. They should operationalize the component of their mission which relates to patient satisfaction and their intentions in meeting patients' needs and wants.

A final question still exists at the end of this study: what if we ignore the changes in satisfaction and in expectations? If we ignore these changes it will be much more difficult to initiate ways to meet patients' needs and wants. It will also be practically impossible to preview and create preventive measures to accomplish compliance from the patients regarding treatment.

References

Beach, L. *et al.* (1976), Developing and testing a decision aid for birth planning decision, *Organizational Behavior and Human Performance*, 5, 99-116.

Borcherding, K.*et al.* (1989), *Comparisons of weighting judgments in multiattribute utility measurement*, Darmstadt, Institut fuer Psychologie Technische Hochschule Darmstadt.

Christensen-Szalanski, J. (1984), Discount functions and the measurement of patients' values. Women's decisions during childbirth, *Medical Decision Making*, 4, 47-58.

Donabedian, A. (1980), *Explorations in quality assessment and monitoring. Volume I: The definition of quality and approaches to its assessment*, Ann Arbor, MI, Health Administration Press.

Drummond, M. *et al.* (1987), Methods for the economic evaluation of health care programmes, Oxford, Oxford University Press.

Edwards, W. (1977), How to use multi-attribute utility measurement for social decision-making, *IEEE Transactions on Systems, Man and Cybernetics*, 326-340.

Farquhar, P. (1984), Utility assessment methods, *Management Science,* 30, 283.

Fishbein, M. and Ajzen, I. (1975), *Belief, attitude, intention and behavior: An introduction to theory and research*, Reading, Mass: Addison-Wesley.

Fishburn, P. (1977), Multiplicative utilities in expected utility theory, *in* Bell, D., Keeney, R. and Raiffa, H. (eds),*Conflicting Objectives in Decisions,* New York, Wiley, 172-194.

Froberg, D. and Kane, R (1989), Methodology for measuring health-state preferences, *Journal of Clinical Epidemiology,* 42, 4, 345-354.

Gustafson, D. and Johnson, S. (1988), *The Effectiveness of Emergency Medical Services and their Categorization for Patients with Major Trauma and Cardiac Conditions.* Final Report submitted to the National Center for Health Services Research and Development.

Gustafson, D. *et al.* (1990), The quality assessment index (QAI) for measuring nursing home quality, *Health Services Research*, 25, 97-127.

Hammond, K.*et al.* (1980), *Human Judgment and Decision Making. Theories, Methods, and Procedures.* New York, NY, Praeger.

Hughes, S.*et al.* (1988), Impact of long-term home care on mortality, functional status, and unmet needs, *Health Services Research*, 23, 269-294.

Hulka, B.*et al.* (1971), Satisfaction with medical care in a low income population, *Journal of Chronic Diseases* , 24, 661-673.

Hulka, B.*et al.* (1975), Correlates of satisfaction and dissatisfaction with medical care: A community perspective, *Medical Care*, 13, 648-658.

Kaplan, R. and Ernst, J. (1983), Do category rating scales produce biased preference weights for a health index? *Medical Care,* 21, 193-207.

Keeney, R. and Raiffa, H. (1976), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* New York, NY, Wiley.

Kincaid, W. (1981), The future of quality assurance in patient care, *in Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care* , Washington, DC.

Krischer, J. (1976), Indexes of severity: Underlying concepts, *Health Services Research*, 143-157.

Linder-Pelz, S. (1982), Social psychological determinants of patient satisfaction: A test of five hypotheses, *Social Science and Medicine*, 16, 583-589.

McNeil, B.*et al.* (1978), Fallacy of the five-year survival in lung cancer, *The New England Journal of Medicine*, 299, 1397-1401.

McNeil, B.*et al.* (1981), Tradeoffs between quality and quantity of life in laryngeal cancer, *The New England Journal of Medicine*, 305, 982-987.

Meterko, M., Nelson, E. and Rubin, H. (1990), Patient judgments of hospital quality. Report of a pilot study, *Medical Care*, 28.

Moles, A. (1982), A new medical oath. *The New England Journal of Medicine*, 307, 761.

Nelson, E.,*et al.* (1989), The patient judgment system: Reliability and validity, *Quality Review Bulletin*, 6, 185-191.

Nunnally, J. (1978), *Psychometric theory*, New York, NY, McGraw-Hill.

Patrick, D. *et al.* (1973), Methods for measuring levels of well-being for a health status index, *Health Services Research*, 228-245.

Pliskin, J.*et al.* (1980), Utility functions for life years and health status, *Operations Research*, 18, 206-224.

Torrance, G. *et al.* (1982), Application of multi-attribute utility theory to measure social preferences for health state, *Operations Research*, 30, 1043-1069.

Torrance, G. (1986), Measurement of health state utilities for economic appraisal, *Journal of Health Economics*, 5, 1-30.

Von Neumann, J. and Morgenstern, 0. (1944), *Theory of Games and Economic Behavior,* Princeton, Princeton University Press.

Von Winterfeldt, D. and Edwards, W. (1986), *Decision Analysis and Behavioral Research.* Cambridge, MA, Cambridge University Press.

Vuori, H. (1982), *Quality Assurance of Health Services: Concepts and Methodology*, Copenhagen: Regional Office for Europe, World Health Organization.

Ware, J. and Snyder, M. (1975), Dimensions of patient attitudes regarding doctors and medical care services, *Medical Care*, 13, 669-682.

Ware, J.*et al.* (1983), Defining and measuring patient satisfaction with medical care, *Evaluation and Program Planning*, 6, 247-263.

Wennberg, J. (1987), The medical care outcome problem: An agenda for actions. *Health Affairs*, 6, 6-32.