

Reliability of a hospital utilization review method in Turkey

SIDIKA KAYA¹, YURDAGÜL ERDEM², SAADET DOĞRUSÖZ³ AND NESRİN HALICI⁴

¹School of Health Administration, Hacettepe University, 06100 Ankara, ²School of Nursing, Fatih University, 06510 Ankara, ³Tuzluca Health Center, Ankara and ⁴Department of Infectious Diseases, Ankara Hospital, Ankara, Turkey

Abstract

Objective. To determine whether the Appropriateness Evaluation Protocol (AEP) is reliable in Turkey.

Methods. Three reviewers, two physicians and one nurse each reviewed 196 patient-days concurrently by using the AEP at three hospitals, two of which were teaching hospitals. Inter-reviewer reliability was assessed both for all cases reviewed (overall agreement), and for only those judged inappropriate by at least one reviewer (specific agreement). In addition, overall agreement between pairs of reviewers was evaluated by the Kappa statistic.

Results. The overall agreement between pairs of reviewers was very high: 93.4–95.9%, and it was similar between all pairs. The level of overall agreement was highly statistically significant: $k = 0.725–0.833$, $P < 0.001$. The specific agreement rates ranged from a low of 61.8% to a high of 75%.

Conclusions. These results show, for the first time, that the AEP method is reliable in Turkey.

Keywords: Appropriateness Evaluation Protocol (AEP), reliability, Turkey, utilization review

Introduction

Utilization review is review of the patient's medical record through application of defined criteria and/or expert opinion. The purpose of utilization review is to assess the efficiency of the health care process and the appropriateness of decision making related to the site of care, its frequency, and its duration. The goal of utilization review and of a utilization management program is to identify and reduce unnecessary or inappropriate hospital use while maintaining access to needed utilization. Inappropriate hospital utilization can be defined as utilization which is not suitable to the patient's medical need [1].

Although utilization review could be used to detect under-utilization – that is the failure to provide necessary services – as well as to identify over-utilization, typically it is restricted to the latter. Over-utilization has been the focus of most utilization reviews for several reasons. The first is concerns about rising health care costs, and the second one is the growing recognition in the past two decades of iatrogenic risk [2]. It is also related to separation of the patient from family, friends, and work. In addition, resources currently used to provide unnecessary services will be made available to fulfil unmet needs by reducing over-utilization.

To formulate a policy for decreasing over-utilization, a valid

and reliable method for determining unnecessary hospital days of care is vital. Although several methods have been used in the past to identify inappropriate hospital use, fundamental methodological problems with the measurement techniques used – notably poor inter-reviewer reliability, bias, lack of comprehensiveness, and sampling difficulties – limited the usefulness of those methods. Particularly critical in the poor performance of the instruments was the reliance on subjective, implicit criteria. Awareness of the shortcomings of these existing approaches for determining the appropriateness of hospital days of care was the impetus for developing an instrument for assessing hospital use, the Appropriateness Evaluation Protocol (AEP). It was developed by the Boston University Health Care Research Unit and met the demands of the health care system in providing useful, objective, generic criteria for assessing the appropriateness (medical necessity) of hospitalization in an acute care facility. In several trials conducted by researchers in the United States its validity and reliability have been confirmed. In the ensuing decade, the AEP gained widespread acceptance in screening cases, concurrently for physician advisor review and retrospectively for profiling provider practice patterns in a variety of institutional settings, especially in the USA [1–12].

The AEP was found to be of similar reliability in Israel [13], and was used to measure and understand the reasons

Address correspondence to Sidika Kaya. Tel: (+90) 312 311 55 06. Fax: (+90) 312 309 36 25 or 90 312 311 93 17. E-mail: sjyl-e@tr-net.net.tr

for inappropriate hospital utilization [14,15]. The AEP (with variable modifications) was also used as a review instrument by researchers in Portugal [16], Spain [17], Italy [18], Switzerland [19], and the UK [20]. Since publication of the method, no studies reporting the reliability of the AEP in Turkey have been reported.

Turkey spent 4.06% of its gross national product on health in 1994 [21], and the budget of the Ministry of Health as a percentage of the national budget was 3.7% in 1995 [22]. From 1988 to 1991, hospital costs rose from 27.8% of total health care expenditures to 35.1% [23]. Because of the magnitude of hospital costs and their rapid increase, it is important to prevent unnecessary expenditure on hospital services. However, today no utilization review program is implemented in hospitals in Turkey.

The objective of this study was to determine whether the AEP, which is reliable in the USA and Israel, is also reliable in Turkey.

Methods

This study was conducted in three hospitals in Ankara, Turkey. One of the hospitals was a university hospital and the others were government general hospitals, one of which was a large teaching hospital. The study population comprised the patients hospitalized in the internal medicine, general surgery, and gynaecology departments of these three hospitals. The unit of evaluation was a single hospitalization day of a patient who was admitted at least 1 day before the evaluated day (the day of discharge was excluded). All patients hospitalized in each department on one randomly chosen day in 1996 were reviewed concurrently. 'Concurrent' review consisted of evaluation of the previous day's appropriateness; reviewers were allowed to obtain data directly from hospital staff and patients as well as information in the medical record as of the day following the review date [3]. Since the objective of this study was to compare the reviewers' judgements on the patient-days, and not to search for absolute hospital figures on misutilization, only one day in each department was reviewed. In this way, 196 patient-days were reviewed.

Inappropriate utilization was measured through application of the AEP developed by Gertman and Restuccia [3]. The AEP provides both a day-of-admission review and a day-of-stay review. In this study, only the day-of-stay was reviewed. We intend to focus on the day-of-admission review in a separate study.

The AEP contains 27 explicit patient-day criteria divided into three groups: medical services, nursing/life support services, and patient condition factors. The first two criteria groups consist of services that were judged by the developers to be routinely available only at an acute hospital level of care. The third group includes factors indicating that the patient's condition was so unstable that immediate availability of acute hospital services was thought to be required. A patient day is considered appropriate if any one of these services or conditions is present. Overrides by the reviewer are, nevertheless, allowed when the criteria do not capture

adequately a patient's situation. For example, a service or factor not included in the criteria may have necessitated the hospitalization, or a patient meeting one of the criteria may not have needed acute level hospitalization. The AEP also contains a list of reasons and responsibilities for inappropriateness to assist in identifying the causes of unnecessary utilization, and to provide guidance on potential interventions to reduce it.

The original AEP was translated into Turkish by an internist and a surgeon separately. Their translations were almost the same, with a few exceptions in wording. They compared their translations and reached an agreement. Then, the original AEP and its Turkish translation were compared by another surgeon. Since he concluded that all items in the two versions had the same meaning, the Turkish translation was accepted as valid.

The face validity of the AEP approach was assessed through critical review of the instrument by 10 internists, 14 surgeons, and six gynaecologists responsible for decisions regarding discharge. These physicians were staff members of two government training hospitals, one university hospital, and a foundation hospital in Ankara. The translated AEP was given to them by the first author. They were requested to examine each criterion and specify if any one should be modified to Turkish medical practice. They did not work as a group; each reported his/her views separately.

Ten percent of these physicians reported that the 'Intramuscular and/or subcutaneous injections at least twice daily' criterion in the AEP would justify the patient's being in an acute care hospital in Turkey. Forty-one per cent of the physicians stated that the 'thoracentesis or paracentesis that day' criterion would justify inpatient care in Turkey. Other physicians reported that these two services would be appropriately provided on an outpatient basis. At least 70% of all physicians felt that the other 25 criteria would result in valid decisions related to the appropriateness of hospital utilization in Turkey. Even though the agreement among the physicians regarding the injections and the thoracentesis or paracentesis criteria was not very strong, these criteria were not deleted from the original AEP. This decision was made because some physicians accepted them as valid, and because it was considered that deletion of criteria should be done if the consequences of statistical analysis of a database containing a large number of days of stay showed that these criteria had poor discriminatory power or were never applied alone. Another reason for not deleting these criteria was to compare the reliability of the AEP in Turkey with that found in other countries. Accordingly, the translation of the original AEP was used in this study.

The three reviewers were a nurse with a PhD and two physicians, the authors of the article. Before the reviews were conducted, the first author trained the reviewers to use the protocol by using the AEP reviewers' manual. The rationale underlying the AEP, application of the AEP criteria, consideration of override, determination of appropriateness, assignment of reasons for inappropriateness, uses of the AEP, and instructions for the data collection form were explained in 1-week training sessions. Then, a sample of 20 patients

Table 1 Inter-reviewer agreement on necessity for a patient-day

Decision of reviewer 2	Decision of reviewer 1	
	Necessary	Unnecessary
Necessary	a	b
Unnecessary	c	d

$$\text{Overall agreement} = \frac{a+d}{a+b+c+d} \times 100$$

$$\text{Specific agreement} = \frac{d}{b+c+d} \times 100$$

was selected from various hospitals and reviewed by all reviewers separately. The decisions of the reviewers were compared and discussed with an emphasis on correct application of the override option. After a base-line AEP competence was established, each reviewer working alone reviewed the same 196 patient days.

Inter-reviewer reliability was tested by calculating 95% confidence intervals for the levels of 'overall agreement' and of 'specific agreement' for each pair of reviewers [3,13]. The overall agreement (Table 1) is calculated (%) by dividing the number of days on the necessity of which both the reviewers agreed (a+d) by all the days reviewed (a+b+c+d). The specific agreement is calculated by dividing the number of days found unnecessary by the two reviewers (d) by the number of days found unnecessary by at least one of them (b+c+d). In addition, overall agreement between pairs of reviewers was evaluated by the Kappa statistic, a measure of agreement that is corrected for chance [24]. This measure has become a standard in this field of research. Since one of the objectives of the survey was to study whether a nurse is reliable in using the protocol, the agreement between the nurse and each of the physicians was compared to the agreement between the two physicians.

Results

One hundred and ninety-six patient-days for the 196 patients were examined: 32.2% of them from the university hospital, 46.9% from the government teaching hospital and 20.9% from the other government hospital; 38.8% of the patients were hospitalized in internal medicine, 49% in surgery, and 12.2% in gynaecology. Their mean age was 50.9 years. No statistically significant differences were found among the three hospitals in the distribution of patients by department, age, and marital status, but the percentage of female patients in the university hospital was higher than in the other hospitals ($P < 0.05$).

The percentages of the noted criteria items in all cases reviewed by each reviewer were calculated. The average of the percentages of the three reviewers for each item is shown

in Table 2. The most frequently noted criteria items in the protocol were in the nursing/life support services group. This finding is parallel with that reported by Gertman and Restuccia [3]. In this group, all criteria items were noted by reviewers at least once in judging a day appropriate. However, the most frequently noted item (73.5%) was 'close medical monitoring by nurse at least three times daily, under doctor's orders'. Among medical services criteria, the most frequently noted item (12.4%) was 'scheduled for procedure in operating room the next day, requiring preoperative consultation or evaluation,' while the criteria 'angiography that day', 'thoracentesis or paracentesis that day', and 'invasive central nervous system diagnostic procedure (e.g., lumbar puncture, cisternal tap, ventricular tap, pneumoencephalography) that day' were not met in any reviews. The most frequent patient condition criterion, 'fever at least 38.3°C rectally (at least 37.7°C orally) if the patient was admitted for reasons other than fever', was met in 10% of the reviews, but the coma criterion was not met in any reviews. Almost half of the patients met three or more criteria.

As shown in Table 3, the overall agreement between pairs of reviewers was very high: 93.4–95.9%, and it was similar between all pairs. The level of overall agreement was highly statistically significant ($k = 0.725-0.833$; $P < 0.001$). The specific agreement rates ranged from a low of 61.8% to a high of 75%.

In internal medicine departments, the level of overall agreement between pairs of reviewers was lower than that found in surgery and gynaecology. In internal medicine, overall agreement between physicians was 94.7% while the levels of overall agreement for each nurse-physician pair were 88.1% and 90.7%. These percentages were 95.8, 96.8, 96.8 in surgery, and 100.0, 95.8, 95.8 in gynaecology, respectively. The level of specific agreement between physicians was lower in surgery (66.6%) than that found in internal medicine (73.3%) and in gynaecology (100.0). On the other hand, the level of specific agreement for each physician-nurse pair was lowest in internal medicine (50.0–56.2%). In surgery, this level was the same for each physician-nurse pair (72.7%), as in gynaecology (80.0%).

The override option was used in nine cases (4.6%) by one of the physicians, in 10 cases (5.1%) by the other physician, and in six cases (3.1%) by the nurse. On average, it was used in 4.3% of all decisions to make a day inappropriate after one or more of the objective criteria had been met. In no case did the reviewers make opposing override decisions. In half of the cases for which the decision made using the criteria alone was overridden by at least one reviewer, two or all reviewers agreed to use the override. Before deciding to use the override option, all reviewers obtained additional information from another source (in most cases this was the nurse) in addition to the patient's records. The assessments of appropriateness that are reported reflect final judgements based on criteria with overrides.

In a case where the only criterion met was the injections criterion, all reviewers overrode the decision made using that criterion alone. 'The criterion for thoracentesis or paracentesis was not met that day'. Therefore, if these two criteria had

Table 2 Percentage distribution of AEP criteria items met in all patient-days reviewed

Criteria Items	% ¹
A. Medical services	
1. Procedure in operating room that day	11.0
2. Scheduled for procedure in operating room the next day, requiring preoperative consultation or evaluation	12.4
3. Cardiac catheterization that day	0.2
4. Angiography that day	0
5. Biopsy of internal organ that day	2.0
6. Thoracentesis or paracentesis that day	0
7. Invasive central nervous system diagnostic procedure (e.g. lumbar puncture, cisternal tap, ventricular tap, pneumoencephalography) that day	0
8. Any test requiring strict dietary control, for the duration of the diet	7.1
9. New or experimental treatment requiring frequent dose adjustments under direct medical supervision	1.2
10. Close medical monitoring by a doctor at least three times daily	2.7
11. Postoperative day for any procedure covered in number 1 or 3–7 above	3.7
B. Nursing/life support services	
1. Respiratory care – intermittent or continuous respirator use and/or inhalation therapy (with chest PT, IPPB) at least three times a day	7.8
2. Parenteral therapy – intermittent or continuous i.v. fluid with any supplementation (electrolytes, protein, medications)	43.4
3. Continuous vital sign monitoring – at least every 30 min, for at least 4 hours	5.6
4. Intramuscular and/or subcutaneous injections at least twice daily	30.3
5. Intake and output measurement	16.7
6. Major surgical wound and drainage care (chest tubes, T-tubes, hemovac, Penrose drains)	14.3
7. Close medical monitoring by nurse at least three times daily, under doctor's orders	73.5
C. Patient condition factors	
Within 24 hours before day of review	
1. Inability to void or move bowels (past 24 hours) not attributable to neurologic disorder	3.6
Within 48 hours before day of review	
2. Transfusion due to blood loss	3.1
3. Ventricular fibrillation or ECG evidence of acute ischemia, as stated in progress note or in ECG report	2.9
4. Fever at least 38.3°C rectally (at least 37.7°C orally), if patient was admitted for reasons other than fever	10.0
5. Coma – unresponsiveness for at least 1 hour	0
6. Acute confusional state, not due to alcohol withdrawal	1.0
7. Acute hematologic disorders, significant neutropaenia, anaemia, thrombocytopenia, leukocytosis, erythrocytosis, or thrombocytosis yielding signs or symptoms	7.1
8. Progressive acute neurologic difficulties	0.7
Within 14 days before day of review	
9. Occurrence of a documented, new acute myocardial infarction or cerebrovascular accident (stroke)	2.0

¹ Average percentages of three reviewers

been deleted in accordance with the views of the physicians assessing the face validity of the AEP, the level of agreement between the pairs of reviewers would not have been different from the levels reported above.

When a day was found to be inappropriate at the acute hospital level, the reviewers identified the probable cause of inappropriateness with the assistance of the AEP reasons list developed in the USA [7] and adapted for use in Turkey. All of the adaptations were related to the discharge planning and alternate facility. Discharge planning – that is planning that begins when a patient is admitted for treatment and is aimed

at providing the patient with the proper segments of the health care continuum [25] – is not carried out in hospitals in Turkey. Therefore, the reasons ‘discharge planned, but no orders written’ and ‘failure to initiate/execute timely hospital discharge planning’ were deleted. The reason ‘patient from an unhealthy environment is kept in hospital until either that environment becomes acceptable or an alternate facility is found’ was modified to ‘patient from an unhealthy environment is kept in hospital until that environment becomes acceptable’. Since there are no nursing homes in the Turkish health care system, the only alternate facility to an acute care

Table 3 Agreement between the three pairs of reviewers

Reviewer Pair 1 ¹ Physician 1				Reviewer Pair 2 ² Nurse				Reviewer Pair 3 ³ Nurse			
Physician 2	+	-	Total	Physician 2	+	-	Total	Physician 1	+	-	Total
+	164	3	167	+	162	5	167	+	164	5	169
-	5	24	29	-	8	21	29	-	6	21	27
Total	169	27	196	Total	170	26	196	Total	170	26	196

¹ General agreement = 95.9%, 95% confidence interval = 93.1–98.7%; $\kappa = 0.833$, $P < 0.001$; specific agreement = 75.0%, 95% confidence interval = 60.0–90.0%.

² General agreement = 93.4%, 95% confidence interval = 89.9–96.9%; $\kappa = 0.725$, $P < 0.001$; specific agreement = 61.8%, 95% confidence interval = 45.5–78.1%.

³ General agreement = 94.4%, 95% confidence interval = 91.3–97.5%; $\kappa = 0.760$, $P < 0.001$; specific agreement = 65.6%, 95% confidence interval = 49.2–82.1%.

+ Necessary day; - unnecessary day.

hospital is a chronic care hospital. Hence, the reason 'patient/family rejection of available space at appropriate alternate facility' was modified to 'patient/family unwillingness to be hospitalized in a chronic care hospital'. The reason 'patient is convalescing from an illness, and it is anticipated that there would be less than a 72-hour stay in an alternate facility' was deleted. No reason was added to the original reasons for inappropriateness.

Of the total number of inappropriate days detected by the three reviewers, 44% were judged to have occurred among patients still requiring hospitalization. For this group of patients, the most frequent reason cited was 'problem in hospital scheduling of tests or non-operative procedure', and it was followed by the reasons 'problem in hospital scheduling of operative procedure' and 'delay in receiving results of diagnostic test or consultation needed to direct further evaluation/treatment'. For the 56% of the inappropriate days occurring among patients who could have been discharged, the predominant reason for inappropriateness was 'physician's medical management of patient is overly conservative'. This reason accounted for 46% of all reasons.

On average, for 26% of the days, reviewers used only written information from the patient's files, for 62% they asked a nurse, the physician responsible, the patient or his/her family for further information, and for 12% of the days they obtained information from more than two sources to complete the protocol. In cases where the reviewers disagreed about the appropriateness of the days, the percentages for not getting additional information from the same sources was 50% for the two physicians, 27% for one physician and the nurse, and 69% for the other physician and the nurse. For days agreed to be necessary or unnecessary, these percentages were 42, 40, and 46 respectively. In the case of 28% of the unnecessary days, reviewers used only the patient's files, while for 60% they also asked a nurse or patient, and for 12% they gathered information from more than two sources to identify the probable cause of inappropriateness.

Discussion

The objective of this study was to determine whether the AEP is reliable in Turkey. The level of overall agreement – the rate of agreement for all cases reviewed, whether judged appropriate or inappropriate – ranged from 93.4% to 95.9%. The Kappa statistic was applied to determine the extent to which these agreements differed from agreements occurring by random chance. Pair-wise comparisons with Kappa values in the 0.725–0.833 range showed 'substantial' to 'almost perfect' agreement [26]. All Kappa levels were statistically significant ($P < 0.001$), indicating that agreement was far higher than expected by chance. The level of overall agreement is similar to that reported by Gertman and Restuccia [3], and Rishpon *et al.* [13].

Of all cases judged inappropriate by at least one of the reviewers, both physician reviewers agreed on which days were inappropriate in 75% of these cases, while the nurse reviewer agreed with one physician reviewer in 61.8% and with the other in 65.6% of such cases. The level of specific agreement is also similar to that reported by Gertman and Restuccia [3] but higher than that reported by Rishpon *et al.* [13].

Although the reviewers were less experienced in evaluating the necessity of patient days and they usually found the quality of medical records to be low, the reliability of the AEP is similar to that found in the USA.

The agreement between the two physicians was similar to that between each one of them and the nurse. This finding is consistent with those already reported in two other investigations [3, 13]. It seems that in Turkey, a nurse with a PhD can be a reliable reviewer of the necessity of patient-days by the AEP. However, it should be noted that the inference cannot be made that the average nurse would be able to use the AEP equally well. Future research should address the issue of agreement between nurse reviewers who have different levels of education. In this study only one nurse reviewer with a PhD used the AEP. Another study

comparing agreement between two nurses with PhDs has commenced in two hospitals in Ankara, and one author of the current study, Dr Kaya, is participating in this research. Most of the physicians contacted by the authors reacted positively to the idea that utilization review could be done by a nurse with a PhD.

Even though the study sample was not selected to be representative of the patient populations for the entire year 1996 at the hospitals being studied, a descriptive finding of this study was that 18.9% of patient-days were found to be inappropriate on the basis of the objective criteria plus subjective reviewer judgements. Reviewers used other sources of information in addition to the medical records to complete the protocol in almost 75% of the reviews. So, perhaps it is desirable to avoid reliance on the quality of medical records in Turkey. Ramos-Cuadra *et al.* documented an association between low completeness of the medical record and greater frequency of inappropriate patient-days [27]. If reviewers had used only medical records, they might have found more inappropriate patient-days.

To maintain quality and access to hospital services, it is important for hospital administrators to focus on unnecessary or medically inappropriate services. The AEP is a widely accepted method to use for this purpose. This study shows, for the first time, that the AEP method is reliable in Turkey.

References

1. Payne S. M. C. Identifying and managing inappropriate hospital utilization: A policy synthesis. *Health Services Res.* 1987; **22**: 709.
2. Restuccia J. D. The evolution of hospital utilization review methods in the United States. *Int. J. Qual. Health Care* 1995; **7**: 253.
3. Gertman P. M., Restuccia J. D. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Med. Care* 1981; **19**: 855.
4. Restuccia J. D., Kreger B. E., Payne S. M. C. *et al.* Factors affecting appropriateness of hospital use in Massachusetts. *Health Care Finan. Rev.* 1986; **8**: 47.
5. Restuccia J. D., Payne S. M. C., Lenhart G. *et al.* Assessing the appropriateness of hospital utilization to improve efficiency and competitive position. *Health Care Manage. Rev.* 1987; **12**: 17.
6. Restuccia J. D., Gertman P. M., Dayno S. J. *et al.* A comparative analysis of appropriateness of hospital use. *Health Affairs* 1984; **3**: 130.
7. Payne S. M. C., Ash A., Restuccia J. D. The role of feedback in reducing medically unnecessary hospital use. *Med. Care* 1991; **29** (Suppl.): AS91.
8. Kreger B. E., Restuccia J. D. Assessing the need to hospitalize children: pediatric appropriateness evaluation protocol. *Pediatrics* 1989; **84**: 242.
9. Strumwasser I., Paranjpe N. V., Ronis D. L. *et al.* Reliability and validity of utilization review criteria: appropriateness evaluation protocol, standardized medreview instrument, and intensity-severity-discharge criteria. *Med. Care* 1990; **28**: 95.
10. Ludke R. L., Wakefield D. S., Booth B. M. *et al.* Nonacute utilization of department of veterans affairs inpatient services. *Med. Care* 1991; **29** (Suppl.): AS29.
11. Siu A. L., Sonnenberg F. A., Manning W. G. *et al.* Inappropriate use of hospitals in a randomized trial of health insurance plans. *N. Engl. J. Med.* 1986; **315**: 1259.
12. Wakefield D. S., Pfaller M. A., Hammons G. T., Massanari R. M. Use of the appropriateness evaluation protocol for estimating the incremental costs associated with nosocomial infections. *Med. Care* 1987; **25**: 481.
13. Rishpon S., Lubacsh S., Epstein L. M. Reliability of a method of determining the necessity for hospitalization days in Israel. *Med. Care* 1986; **24**: 279.
14. Rishpon S., Epstein L. M., Rennert H. Unnecessary hospitalization days: rates in two general hospitals in Israel. *Israel J. Med. Sci.* 1989; **25**: 392.
15. Paldi Y., Porath A., Friedman L., Mozes B. Appropriateness of hospital use: a comparative study. *Int. J. Qual. Health Care* 1995; **7**: 261.
16. Bentes M., Gonsalves M. D. L., Santos M., Pina E. Design and development of a utilization review program in Portugal. *Int. J. Qual. Health Care* 1995; **7**: 201.
17. Lorenzo S., Sunol R. An overview of Spanish studies on appropriateness of hospital use. *Int. J. Qual. Health Care* 1995; **7**: 213.
18. Fellin G., Apolone G., Tampieri A. *et al.* Appropriateness of hospital use: an overview of Italian studies. *Int. J. Qual. Health Care* 1995; **7**: 219.
19. Santos-Eggimann B., Paccaud F., Blane T. Medical appropriateness of hospital utilization: an overview of the Swiss experience. *Int. J. Qual. Health Care* 1995; **7**: 227.
20. O'Neill D., Pearson M. Appropriateness of hospital use in the United Kingdom: a review of activity in the field. *Int. J. Qual. Health Care* 1995; **7**: 239.
21. Sağlık Bakanlığı Tedavi Hizmetleri Genel Müdürlüğü, Statistical Yearbook of Hospitals 1995 (in Turkish). Yayın No: 580, Ankara, 1996.
22. Republic of Turkey Ministry of Health, Health Statistics 1994. Publication No. 579, 1995.
23. T. C. Sağlık Bakanlığı Sağlık Projesi Genel Koordinatörlüğü, Financing the Health Sector in Turkey (in Turkish) Ankara, 1993.
24. Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Measure* 1960; **20**: 37.
25. Timmreck T. C. Dictionary of Health Services Management. USA: National Health Publishing, 1987.
26. Landis J. R., Koch G. G. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159.
27. Ramos-Cuadra A., Marion-Buen J., Garcia-Martin M. *et al.* Use of the appropriateness evaluation protocol: the role of medical record quality. *Int. J. Qual. Health Care* 1995; **7**: 267.

Received in revised form 26 June 1997