International Journal for Quality in Health Care 2001; **Volume 13, Number 4**: pp. 289–299

Screening inappropriate hospital days on the basis of routinely available data

PATRICIA HALFON[1] AND YVES EGGLI[2]

[1]The Planning and Evaluation Office, Centre Hospitalier et Universitaire Vaudois Lausanne, and [2]The Institut d'Économie et de Management de la Santé, University of Lausanne, Lausanne, Switzerland

Abstract

Objective. The systematic use of regular hospital utilization reviews has proved costly, particularly in countries with short average lengths of stay. This study examines the performance of three tests based on routinely collected data when screening inappropriate hospital days.

Design. The Appropriateness Evaluation Protocol was used to set the gold standard. The first screening test was simply based on the comparison of an observed length of stay with a target value; the second test additionally made allowances for surgical and intensive care procedures while the third added the amount of required nursing workload to these data.

Setting. The neurology and general surgery departments of a Swiss university hospital.

Participants. Every day of care for all inpatients stays was reviewed to assess the appropriateness of hospital use and submitted to the screening algorithm (9000 hospital days).

Main outcome measures. Receiver-operating characteristics curves were compared to optimize the performance of the screening tests. The best test was applied to all units of the hospital and rates of inappropriate days were computed using a Bayesian approach.

Results. The first and the second tests have a sensitivity of 66–80% and a specificity of 66–67%. Nursing workload data yield no significant improvement of the screening test. An unbiased estimate of the rate of inappropriate days may be computed.

Conclusion. The present study provides some evidence that a screening approach is useful, feasible and efficient for detecting inappropriate hospital days.

Keywords: appropriateness review, Bayes theorem, hospitalization, length of stay, nursing process

The majority of hospital utilization reviews use a costly approach based on a case-by-case evaluation of medical records [1–6]. Many of the studies carried out over the last decade have shown that 15–50% proportions of inappropriate days are common, particularly in European countries [7–10]. However, it should also be noted that lengths of stay have been dramatically reduced in most hospitals in the developed world, suggesting a reduction in inappropriate hospital use [11]. The Swiss experience has shown that only 50% of unnecessary days can be avoided without additional resources [12–15]. Moreover, because they do not require intensive treatment or invasive procedures, inappropriate days are generally less expensive. Review methods should consequently focus on areas with a high probability of inappropriate days.

Quality assurance literature suggests that a set of complementary tools may make the approach more efficient. These tools should follow the quality cycle [16]: detecting or surveillance of a potential problem area, measurement, causal analysis and follow-up of corrective procedures (Figure 1). Data requirements increase from one step to the next.

Highly detailed information that is only available at a local level is generally required to measure the prevalence of a problem and determine its cause. Ultimately, only listening to those involved in the delivery of care can provide the data required for a thorough assessment. On the other hand detection and surveillance, suitable for use at regional or national level, should be based on routinely available data [17]. No such valid instrument has yet been documented.

Address reprint requests to P. Halfon, Institut Universitaire de Médecine Sociale et Préventive, rue du Bugnon 17, CH-1005 Lausanne, Switzerland. Email: patricia.halfon@hospvd.ch

The views expressed herein are those of the authors and do not necessarily reflect the views of the authorities concerned.

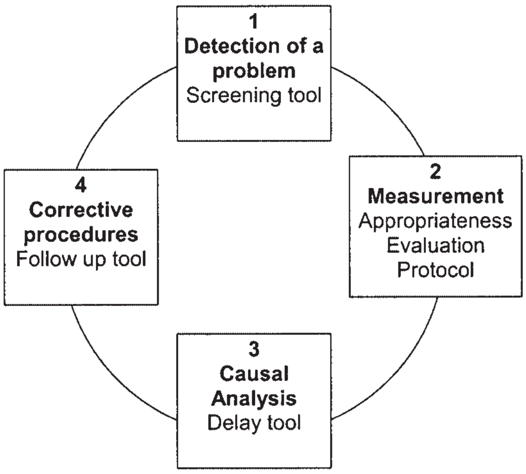

Figure 1 The quality cycle.

To measure the appropriateness of hospital days, many studies first used the Appropriateness Evaluation Protocol (AEP) without a screening tool [1,3,4,6–10,12–15]. Medical records of a sample of hospital days were reviewed on a case-by-case basis according to a set of explicit criteria. All inappropriate days were then analysed with a Delay tool to document the causes [1,18]; potential corrective procedures were finally subjected to a cost–benefit analysis. Unfortunately, many reviews were wasted on cases with low problem probability. The American College of Physicians has recommended that routine case-by-case reviews should be dropped and patterns of care at institutional, regional or national levels should be profiled instead [19].

Since the amount of inappropriate uses across diagnostic categories were found to vary widely, some authors suggest that reviews should target patients in selected high risk groups [20,21]. This approach neither yields an unbiased estimate, however, nor does it allow the adaption of the selection criteria when shifts in inappropriate care occur. Screening methods based on weighted probability sampling have been proposed to address these issues, but they focus on inappropriate admissions rather than days [22–24]. Other authors suggested simply using the mean length of stay as a proxy of inappropriate days but they did not find this method particularly effective, in spite of a standardized hospital case mix [2,25,26].

Reviews conducted in Canton Vaud in the 1990s highlighted the fact that inappropriate admissions were rare (less than 6%), that they were often followed by appropriate days, and that delays predominated in the discharge process, responsible for 52–82% of the inappropriate days [3,4,7,12–15]. Other studies confirmed that inappropriate days occur more frequently at the end of stays [27,28]; and furthermore, that after 1 week more than half the patients requiring acute care at admission no longer needed hospital care [29]. These findings suggest that the distribution of lengths of stay should be the point of interest.

The aim of this paper is to complete the quality cycle by providing a tool to screen inappropriate hospital days, addressing the following issues: (i) what kind of information is required to build an effective detection tool? (ii) is it possible to use only data routinely collected by health authorities throughout the majority of developed countries? (iii) if not, which additional data should be collected?

Three screening tools were developed and compared (Figure 2).

(i) The first one considers any day exceeding a defined length of stay threshold as inappropriate; the threshold is determined by Diagnosis Related Groups (DRGs) which are routinely available through most national databases in developed countries.
(ii) The second screen also accounts for intensive and intermediate care, operation room utilization and deaths. Such a screening test is easy to calculate at hospital level for use by hospital managers.
(iii) The third tool (available in some hospitals only) accounts for nursing workload as well as the conditions cited in the second screen.

Methods

Material used to screen inappropriate days

The study was conducted in the Centre Hospitalier Universitaire Vaudois (CHUV), a Swiss non-profit teaching hospital with 818 beds, 23 000 inpatients per year and an average length of stay of 8.6 days. The CHUV has extensive experience in AEP implementation [3,4,14,15], and a structured executive information system [30].

The population studied included all patients admitted to the neurology department between 1 October 1996 and 31 March 1997, and to one of the two general surgery departments between 1 November 1996 and 30 April 1997. The main purpose of the review in neurology was to assess the effect of corrective measures to decrease discharge delays identified by a previous review [31]; the surgery department was chosen with a view to getting a clear picture of other causes of delay. Zero-time was the day of admission to the hospital; the follow-up period included stays in emergency, intensive care, general surgery and neurology departments, but not stays in other clinical units (< 5% of internal transfers). Each day of the patient's stay was reviewed individually. Days of discharge were excluded from the analysis.

The following data were extracted from the hospital information system: patient identifier, admission and discharge dates, intensive and intermediate care dates, All Patients Diagnosis Related Groups (AP-DRGs, version 12.0) [32], date of procedure in an operating room, discharge status (alive or dead) and nursing workload measured by the direct required time of care in hours per patient days according to the Project Research in Nursing (PRN) method [33]. Detailed nursing activities according to the PRN method and review data were collected in a computerized database during the study; they were later linked to data in the hospital information system by means of a hospitalization identifier.

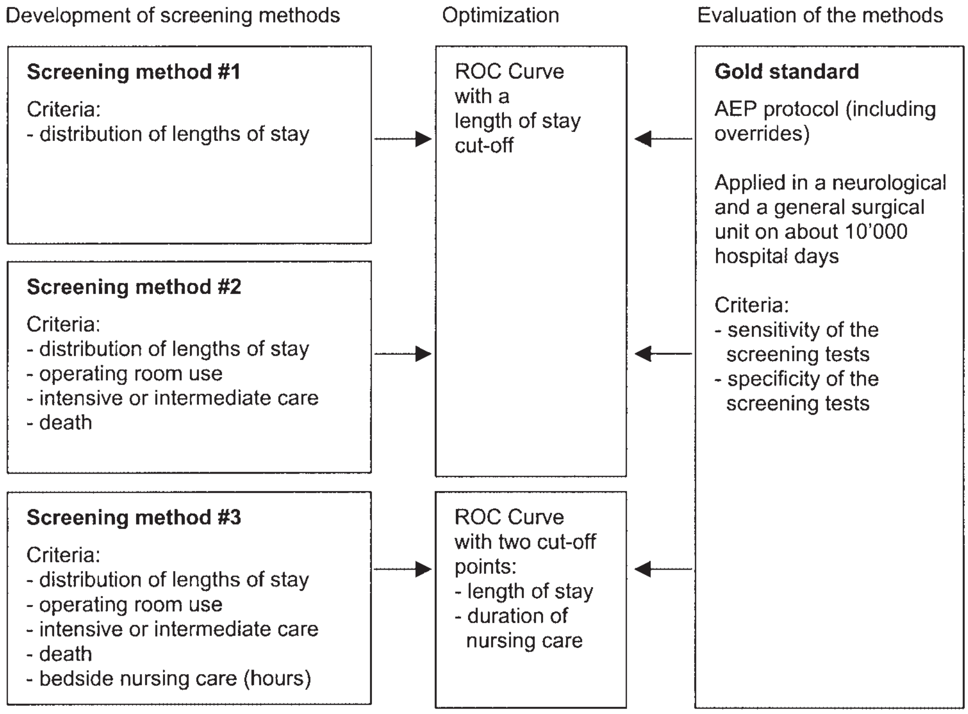

Figure 2 Design and optimization of the three screening methods. The three screening methods were built on *a priori* criteria. The gold standard was established independently on the basis of the AEP. The screens were then optimized in order to maximize the sensitivity and the specificity of the tests, using ROC curves with one (methods #1 and #2) or two (method #3) variables. AEP, Appropriateness Evaluation Protocol; ROC, Receiver Operating Characteristics.

Gold standard

The gold standard was developed independently of the screening tests. The appropriateness of hospital days was based on the first version of AEP criteria (Appendix A) [1]. The criteria justifying an acute hospital day are categorized into three batteries relating to:

(i) Medical services (operating room procedures, arterial angiography, thoracocentesis, etc.);
(ii) Nursing services (vital signs monitoring, respiratory devices, wound nursing, etc.);
(iii) Patient characteristics (acute myocardial infarction, pulse rate, blood pressure, coma, etc.).

To establish a gold standard, the protocol allows overriding of AEP criteria. These override rules, which were set down by a panel of ten physicians and nurses from the hospital (review committee) before the beginning of the review are presented in Appendix B in the form of a decision table [34]. Overrides may also have been used by the reviewer in certain conflicting situations; in these cases override decisions had to be approved by heads of the departments concerned.

As with most utilization reviews conducted in non-profit hospitals in Canton Vaud (Switzerland), this review was carried out concurrently with inpatient stay in order to promote information exchange between hospital staff and reviewer [3,4,12–15].The reviewers, a nurse for surgery and a physician for neurology, read medical and nursing records (including notes, order sheets and lab reports) within two days following admission and then two or three times a week. They could obtain additional information through interviews with staff members whenever necessary. Both reviewers were trained by the second author during practice sessions, then closely supervised during the first two weeks. Particular care was taken to formulate a separate assessment of every day under review; this had to be based on the objective data available for that day only.

The rate of inappropriate days was defined as the number of inappropriate days divided by the total of the reviewed hospital days.

Optimization of screening tests

Since the focus is on the detection of inappropriate days, the observation units were individual days of care. Days with missing screening criteria variables were excluded from the analysis.

The first screening test is based on a simple rule: each day of care beyond an expected value according to its diagnosis group is considered inappropriate. A benchmarking approach was applied to determine length of stay standards for each group of patients, classified by AP-DRGs [35].

The second screening test is based on the same rules, with additional criteria aiming to improve the specificity of the test:

(i) every day in an intensive or intermediate care unit is justified;

(ii) the days before, during and following an operating room procedure are justified;
(iii) the day preceding a death is justified.

In addition to the previous model, the third screening test includes the required duration of nursing care. A day is also considered as justified if the workload exceeds a cut-off value (four nursing hours during the day for instance).

In order to build a receiver-operating characteristics (ROC) curve, different cut-off points were used, ranging from 25 to 200% of the standard values of the length of stay for the two first tests. For the third test several ROC curves were built using different cut-off values of the nursing workload and the day rank.

The best cut-off value (for day rank or nursing care hours) was chosen so as to maximize the expected utility of doing the screening test. Demonstrably this condition is reached where the slope S of the ROC curve equals [36]:

$$S = (C/B)\ (1 - p)/p \quad \text{[Formula 1]}$$

where C is the mean cost of the review per day, B is the mean potential savings per inappropriate day and p is the prevalence of inappropriate days. The C/B ratio was estimated to be 0.14 (Appendix C).

Screening accuracy was assessed by the sensitivity, specificity and likelihood ratios of the test. Sensitivity was measured by the proportion of inappropriate days with a positive test result, specificity by the proportion of appropriate days with a negative test result. The likelihood ratio (Lr_i) of the results of the test (i = positive or negative) is the conditional probability of the result of the test if the day is inappropriate divided by the conditional probability of this result if the day is appropriate. It indicates the extent to which the screening test result will raise or lower the target event's *a priori* probability.

Estimated rate of inappropriate days

Inappropriate days (I) are the result from the union of true positives (TP) and false negatives (FN). The following formula can be used to estimate the rate of inappropriate days [37]:

$$p = r(+)\ p(I\,|\,+) + r(-)\ p(I\,|\,-) \quad \text{[Formula 2]}$$

where r(+) is the rate of positive test results, r(−) is the rate of negative test results and $p\ (I\,|\,+)$ and $p\ (I\,|\,-)$ are *a posteriori* probabilities. These are the probability of being inappropriate in the presence of a positive test result, and in the presence of a negative test result, respectively. Each *a posteriori* probability can be calculated via Bayes theorem which relates it to the *a priori* probability and the Lr_i [38].

The *a priori* probability was estimated by the maximum relative excess of observed days compared with expected days, and a basal rate of 5%:

$$p' = \max\{5\%;\ \textstyle\sum_i p_i\ [(L^a_i - L^s_i)/L^a_i] + 5\%]\} \quad \text{[Formula 3]}$$

where i indexes AP-DRG, L^a is the actual average length of stay, L^s is the standard average length of stay and p is the proportion of patients. The minimum basal rate was set at 5% because published inappropriate days were always higher than 5% [2,6,7].

Some assumptions have to be verified to enable the application of the Bayes' theorem. First, the true state (gold standard) should be established without using the same information as used in the screening test; the present protocol globally meets this condition except for operating room utilization data. Second, the likelihood ratio has to be universal, i.e. independent of the type of unit. Third, the conditional probabilities of the test results must be proven to be independent of prior information [39].

Therefore, any discussion of the results should tackle these questions, particularly when supporting the following assertions:

(i) likelihood ratios are not significantly different in two contrasting care units (general surgery and neurology);
(ii) when the screening test is applied to all care units in the hospital, the estimated rate of inappropriate days is not equal to the *a priori* probability;
(iii) the probability of having a positive or negative test if the day is inappropriate is not linked to *a priori* probability.

Results

During the 6 month study period, 582 patients were admitted to the department of neurology; they totalled 4282 hospital days (day of discharge excluded); there were 581 surgery admissions and 4949 hospital days. The mean nursing care workload per patient day was 3.2 hours (range: 0.8–9.1) in neurology and 3.5 (range: 0.5–15) in surgery.

Gold standard

On the basis of AEP and override criteria, 18.6% of neurology days and 7.7% of surgery days were inappropriate (Table 1). In both units, overrides were used more frequently to remove than to add appropriate days. In neurology, the most frequent cause of justifying days that did not meet explicit criteria was observation to establish a diagnosis, while the most frequent cause to delete a day was the use of restrictive criteria for stroke and transient attack. In surgery, almost 5% of days were judged inappropriate in spite of an explicit criterion related to nursing services, mostly because the required nursing care could have been provided on an outpatient basis. It should be stressed that the frequencies indicated in Table 1 under 'removed days' reflect situations, where the criterion mentioned was the only one justifying a day. For instance, the use of continuous intravenous fluid was frequent in neurology; it was the only criterion for 157 days. Nursing care activities, considered as specific to acute hospitalization were however frequently collected in the daily records of the PRN methods [33]. Thus, only 44 days were considered inappropriate (Appendix B, column V). Figure 3 shows that the proportion of inappropriate days is higher at the end of the stay.

Table 1 Appropriateness of days in the neurology and surgery departments

	Surgery		Neurology	
Criteria	Number of days	%	Number of days	%
AEP justified days	4751	96.0	3661	85.5
Additional days	57	1.1	57	1.3
– difficult investigation	44		13	
– difficult physiotherapy	3		1	
– necessary observation	7		41	
– humanitarian criteria	3		2	
Removed days	−238	−4.8	−232	−5.4
– strict dietary control	−3		−30	
– medical monitoring	−234		−16	
– intravenous fluid	−1		−44	
– i.m. or s.c. injections[1]	0		0	
– stroke or transient attack	0		−142	
Appropriate days	4570	92.3	3486	81.4
Inappropriate days	379	7.7	796	18.6
Total	4949	100.0	4282	100.0

[1] i.m. = intramuscular; s.c. = subcutaneous.

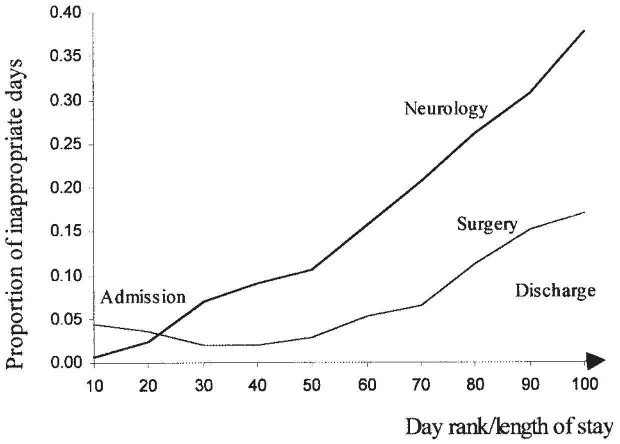

Figure 3 Rates of inappropriate days during stays in the two departments.

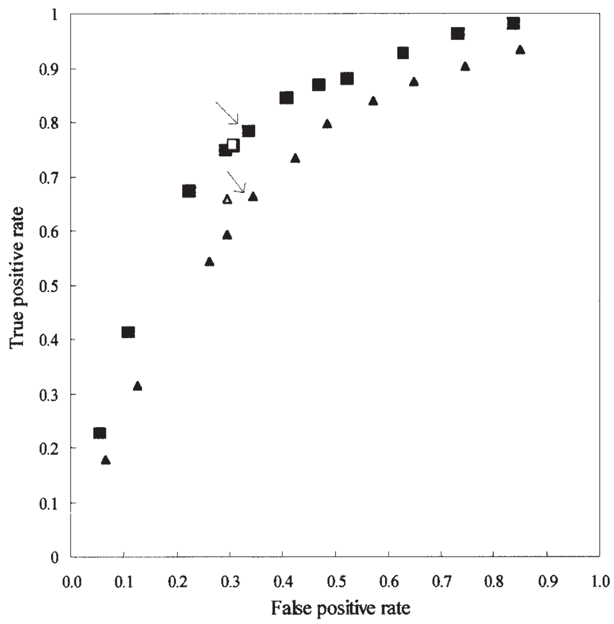

Figure 4 ROC curves over a range of cut-off values of length of stay (LOS). ■ = Neurology; ▲ = Surgery; arrow = optimal cut-off point of the test based only on LOS (0.8*LOS); □ = performance of the second test at a cut-off point of 0.8*LOS in Neurology; △ = performance of the second test at a cut-off point of 0.8*LOS in Surgery.

Screening tests

The ROC curves plot the sensitivity (true positive rate of inappropriate days) against one minus specificity (false positive rate) over a range of cut-off points from 25 to 200% of the expected length of stay (Figure 4). Each point of the curve corresponds to a cut-off value, above which a day is considered inappropriate. The diagonal line represents a test without discriminative ability. As shown by the areas under the ROC curves, the first screening test (based only on the day rank) performed better in neurology than in surgery. In neurology, adding data (especially intensive or intermediate care, operating room, etc.) does not enhance the screen's discriminative ability. In surgery, additional data tended to increase test specificity. The optimal cut-off values of the day rank were 80% of the expected length of stay in both services for the first test and 75% for the second test.

The ROC curve points corresponding to different values of the nursing care hours per day (from 1.5 to 6 hours) are close to the diagonal, showing that this variable has no

Table 2 Comparative performances of the screening tests in neurology

	Gold standard				
	Inappropriate days		Appropriate days		
Test result	Number	Proportion	Number	Proportion	LR_i[1]
Screening test 1					
Inappropriate days	606	606/770 = 0.79[2]	1119	1119/3348 = 0.33	2.42
Appropriate days	164	164/770 = 0.21	2229	2229/3348 = 0.67[3]	0.32
Missing variables	26		138		
Screening test 2					
Inappropriate days	590	590/770 = 0.77	1116	1116/3348 = 0.33	2.30
Appropriate days	180	180/770 = 0.23	2232	2232/3348 = 0.67	0.35
Missing variables	26		138		
Screening test 3					
Inappropriate days	462	462/659 = 0.70	909	909/3006 = 0.30	2.32
Appropriate days	197	197/659 = 0.30	2097	2097/3006 = 0.70	0.43
Missing variables	137		480		

[1] LR_i indicates the likelihood ratio of a positive test (inappropriate day) or a negative test (appropriate day).
[2] Measures the sensitivity of the test.
[3] Measures the specificity of the test.

Table 3 Comparative performances of the screening tests in surgery

	Gold standard				
	Inappropriate days		Appropriate days		
Test result	Number	Proportion	Number	Proportion	LR_i[1]
Screening test 1					
Inappropriate days	248	248/374 = 0.66[2]	1523	1519/4441 = 0.34	1.94
Appropriate days	126	126/374 = 0.34	2918	2918/4441 = 0.66[3]	0.51
Missing variables	5		129		
Screening test 2					
Inappropriate days	259	259/374 = 0.69	1444	1444/4441 = 0.33	2.1
Appropriate days	115	115/374 = 0.32	2997	2997/4441 = 0.67	0.46
Missing variables	5		129		
Screening test 3					
Inappropriate days	234	234/350 = 0.67	1315	1315/4146 = 0.32	2.11
Appropriate days	116	116/350 = 0.31	2831	2831/4146 = 0.68	0.49
Missing variables	29		424		

[1] LR_i indicates the likelihood ratio of a positive test (inappropriate day) or a negative test (appropriate day).
[2] Measures the sensitivity of the test.
[3] Measures the specificity of the test.

discriminative ability when applied to all days; when applied to days beyond 0.75 expected length of stay, a cut-off value of 5 hours' nursing improves test specificity of the test but decreases its sensitivity (data not shown). The performances of the three screening tests are given in Tables 2 and 3.

The following results, suggesting that the Bayesian analysis is not biased, are derived from the application of the second screening test. The likelihood ratios, calculated while pooling the days of care of the two units were 2.3 and 0.4.

Application of the test to data from the same period as the review produces an estimated value of 17.9 in neurology, compared with an actual rate of 18.6, and an estimated value of 8.5 in surgery, which corresponds to an actual rate of 7.7. Moreover, the assumption of the independence of conditional

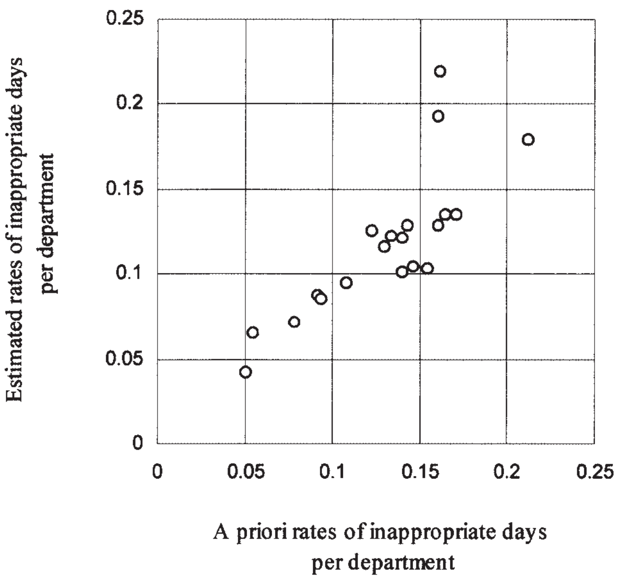

Figure 5 *A priori* and estimated rates of inappropriate days by hospital department.

probabilities of the test and prior information is verified, since the correlation coefficient between these variables computed for each AP-DRG is equal to 0.02.

The estimated rates of inappropriate days in 1997, plotted against *a priori* probabilities for each hospital department (see Figure 5) show that the screening test alters markedly the *a priori* values.

Comments

The first and the second screening tests are able to detect most inappropriate days of care. Nursing workload data were found to be of no value in this regard. On one hand, many inappropriate days correspond to a high nursing workload; patients suffering from a stroke or multiple sclerosis, for example. For most patients, nursing activities related to diagnosis or treatment decreased at the end of the stay, while others activities related to basic patient needs (food, mobility, information) remained relatively constant throughout. On the other hand, the amount of required nursing hours was closely related to the presence of the patient in an intensive or intermediate care unit.

The best screening test was the second one. Because they express severe patient conditions independently of the diagnostic information, the use of additional data – such as intensive care or operating room utilization – allowed for a relatively safe generalization of the test to other specialized care units. These data are collected routinely in our hospital and integrated within the hospital information system. Unfortunately, in most hospitals they are not readily available, so that the cost of the information required rises dramatically. In these cases, the first test, which simply justifies all days until a cut-off point determined by patient diagnosis and surgical interventions, would be the most efficient. Its main advantages are that it is economical and practical: such medical data are readily available in most developed countries.

The method proposed is limited in several ways. The screening test is vulnerable to errors in assigning patients to diagnosis categories. Possible DRG creep (diagnoses are coded in order to maximize the patient's severity of illness) may distort the review process. Inappropriate admissions might on occasion be masked by low average lengths of stay. Inappropriate preoperative days are often missed: in one extensive study 50% of the surgical admissions had no procedures performed within 24 hours of admission, raising questions as to the policies relevant to preadmission testing and intervention planning [40]. In Canton Vaud, the most frequent cause of inappropriate admission in surgery was also an operation not performed in a timely fashion (patients awaiting surgery) [12,14]. Finally, external validity of the method still has to be assessed in other populations.

The gold standard was based on information collected independently from the data used in the screening test. The sensitivity of the test is not correlated with *a priori* information. The likelihood ratios are similar in both care units although they deliver quite different types of care. Estimated rates of inappropriate days are different to *a priori* probabilities. All these observations further confirm consistency of results of the proposed screening test. The test's discriminative power as assessed by the ROC curve is comparable to the prediction model proposed by Ash *et al.* [22] and other models widely used in the outcome measurement field [41].These performances might be improved if an international benchmarking process determining length of stay standards were to be adopted.

The position of the test's optimal cut-off point depends upon utilities determined using Swiss data, so that its external validity may be questionable. The potential savings linked to an inappropriate day and the costs of a review will undoubtedly differ from country to country. However, it is very likely that the utility ratio would not change dramatically. The sensitivity analysis in the published prevalence range shows that the optimal cut-off value, determined on the ROC curve, should remain in the 65–85% bracket of the standard length of stay. This result concords with common sense, moreover, since the first two-thirds of the days of stay are difficult to avoid. Several studies confirm this hypothesis: the greatest proportion of non-acute days of care occurred during the latter third of the stay, and 95% of these late non-acute days did not require continued hospitalization beyond the day reviewed [27]; lastly, the most frequent causes of delays were related to the discharge process [7–10,18,42,43].

The proportion of avoidable days among inappropriate days (50% in our experience) is another reason for concern. The answer depends on the causes of delays and the capacity of the individual hospital to solve relevant problems. Mozes *et al.* [44] showed that pre-set criteria for inpatient care applied during 1 month on a medical unit can decrease the proportion of unjustified days by 53% during the intervention period, compared with a control ward, without adverse impact on hospital death or readmission rates. The effects of utilization review on hospital use in non-experimental design studies are less clear [45].

Except for diagnostic categories, few patient characteristics have been consistently found to be significantly related to inappropriateness [2,20,23,46,47]. No consistent association was found with age, gender, race, income or cost sharing [2, 48,49]. Nor have hospital characteristics been found to be consistently associated with higher rates of inappropriate utilization [2]. The hospital stay characteristics, such as admission day or month, planned surgery, did not predict inappropriate use in multivariate analysis [50]. Studies on admission type (elective versus urgent) are scarce [22]. The most common reasons for inappropriate days identified in several AEP studies are under the responsibility of the physician and/or the hospital [2]. The performance of the predictive model proposed by Ash *et al.* [22] is dramatically improved by adding variables related to physician. All these findings suggest that the proposed screen, based on variation of length of stay, a variable depending mostly on physician practice patterns [51], may be generalizable.

The tests proposed are used to find patterns of inappropriate use and to flag units for further reviews. They only indicate that a problem relative to achieving discharge in a timely manner may exist, bearing in mind that quality assurance activities do not seek to reach a zero defect objective [52]; care evaluation must be assessed in much greater detail at a local level to confirm the problem and identify its causes. Based on limited information, the method is easy to implement in most hospitals. The quality control system could consist of two parts; an automated screening system using universally available hospital data and an audit sample of medical records conducted at the request of the quality manager when the screen signals settings with enhanced probability of inappropriate days. Because screening variables are unknown until the patient is discharged, concurrent review of days screened is not possible. Reviews could be conducted on a concurrent basis on other stays in the flagged units.

Conclusion

The present study provides some evidence that a screening approach is useful and feasible in detecting inappropriate hospital days.

Three screening tests have been applied in two medical departments. The first test identifies excessive hospital days on the basis of 80% of a benchmarking-based standard value stratified by AP-DRGs. Its sensitivity is 79% (neurology) versus 67% (surgery) with a specificity of 67% (both departments). The second test, based on additional data, such as the patient being in intensive, intermediate care units or in the operating room, is the test we recommend if these data are readily available. The third test, based on nursing workload data, yielded no significant improvement, and has been rejected in consequence.

All validation criteria have been met in this study, in particular, the assumptions associated with a Bayesian analysis. The test can be recommended for a focus utilization review approach on units. But, if strategic decisions had to be taken, such as bed planning or unit closure, interhospital comparability should be confirmed in order to ensure a non-biased analysis.

Acknowledgements

The authors are indebted to Brigitte Santos-Eggimann and Guy van Melle for their helpful comments on the manuscript. They thank the reviewers, Céline Hascoët and Veronica Castillo, the medical departments' heads Julien Bogousslavsky (neurology) and Michel Gillet (surgery), Patrick Genoud, Sylvia Hauri and the medical and nursing staffs for their contributions to the study. This work was supported by a grant (#22) from the Performance funds (Public Health Service of Canton de Vaud, Switzerland).

References

1. Gertman PM, Restuccia JD. The Appropriateness Evaluation Protocol: a technique for assessing unnecessary days of hospital care. *Med Care* 1981; **19:** 855–871.
2. Payne SM. Identifying and managing inappropriate hospital utilization: a policy synthesis. *Health Serv Res* 1987; **22:** 709–769.
3. Castillo V, Halfon P, Eggli Y *et al. Hospital Utilization Review in the Neurology Department of Centre Hospitalier Universitaire Vaudois in 1997* Raisons de santé No. 8 (in French). Lausanne: Hospices cantonaux, 1999.
4. Hascoet C, Halfon P, Eggli Y *et al. Hospital Utilization Review in the General Surgery Department of Centre Hospitalier Universitaire Vaudois in 1997* Raisons de santé No. 9 (in French). Lausanne: Hospices cantonaux, 1999.
5. Ash A. The design and analysis of hospital utilization studies. *Int J Qual Health Care* 1995; **7:** 245–252.
6. McDonagh MS, Smith DH, Goddard M. Measuring appropriate use of acute beds: a systematic review of methods and results. *Health Policy* 2000; **53:** 157–184.
7. Santos-Eggimann B, Paccaud F, Blanc T. Medical appropriateness of hospital utilization: an overview of the Swiss experience. *Int J Qual Health Care* 1995; **7:** 227–232.
8. Lorenzo S, Suñol R. An overview of Spanish studies on appropriateness of hospital use. *Int J Qual Health Care* 1995; **7:** 213–218.
9. Bentes M, Gonsalves ML, Santos M, Pina E. Design and development of a utilization review program in Portugal. *Int J Qual Health Care* 1995; **7:** 201–212.
10. Fellin G, Apolone G, Tampieri A *et al.* Appropriateness of hospital use: an overview of Italian studies. *Int J Qual Health care* 1995; **7:** 219–225.
11. *OECD Health Data 2000: a Comparative Analysis of 29 Countries* (database online). Paris: OECD Publications service, 2000.
12. Groupement des Hôpitaux Régionaux Vaudois. *Concurrent Hospital Utilization Reviews in the Hospitals of the Hospital Consortium*

in the Vaudois Region (in French). Prilly: Groupement des Hôpitaux Régionaux Vaudois, 1998.

13. Khalil C, Blanc T, Sidler M. *Inappropriate Hospital Use in Hôpital de zone d'Aigle* Cahier Recherche Documentation IUMSP No. 84 (in French). Lausanne: Hospices cantonaux, 1992.
14. Reutter E, Garin C, Basterrechea L, Livio JJ. *Hospital Utilization Review in the Orthopedic Department of Centre Hospitalier Universitaire Vaudois and in Hôpital orthopédique de la Suisse romande* Cahier Recherche Documentation IUMSP No 118 (in French). Lausanne: Hospices cantonaux, 1992.
15. Christen F, Eggli Y, Nicod P. *Hospital Utilization Review in the Internal Medicine Department of Centre Hospitalier Universitaire Vaudois* Cahier Recherche Documentation IUMSP No 119 (in French). Lausanne: Hospices cantonaux, 1992.
16. Holt P. Quality assessment and improvement. In Australian Council on Healthcare Standards, ed., *The Medical Services: Assessing and Improving Quality.* Ultimo: ACHS, 1994: pp. 63–81.
17. Audet AM, Scott HD. The uniform clinical data set: an evaluation of the proposed national database for Medicare's quality review program. *Ann Intern Med* 1993; **119:** 1209–1213.
18. Selker HP, Beshansky JR, Pauker SG, Kassirer JP. The epidemiology of delays in a teaching hospital: the development and use of a tool that detects unnecessary hospital days. *Med Care* 1989; **27:** 112–129.
19. American College of Physicians. The oversight of medical care: a proposal for reform. *Ann Intern Med* 1994; **120:** 423–431.
20. Payne SM. Targeting utilization review to diagnostic categories. *Qual Rev Bull* 1987; **13:** 394–404.
21. Strumwasser I, Paranjpe N, Ronis DL *et al.* Focused utilization review strategies. *Qual Rev Bull* 1989; **15:** 246–254.
22. Ash A, Schwartz M, Payne SM, Restuccia JD. The self-adapting focused review system: probability sampling of medical records to monitor utilization and quality of care. *Med Care* 1990; **28:** 1025–1039.
23. Saad AN, Andrews RW, Griffith JR. Using claims data to monitor hospital utilization. *Inquiry* 1993; **30:** 447–454.
24. Rosenberg MA, Griffith JR. A management tool for controlling the rate of non-acceptable inpatient hospital claims. *Inquiry* 1999; **36:** 461–470.
25. Houghton A, Bowling A, Jones I, Clarke K. Appropriateness of admission and the last 24 hours of hospital care in medical wards in an east London teaching group hospital. *Int J Qual Health Care* 1996; **8:** 543–553.
26. Hartz AJ, Bade PF, Sigman P *et al.* The evaluation of screening methods to identify medically unnecessary hospital stay for patients with pneumonia. *Int J Qual Health Care* 1996; **8:** 3–11.
27. Booth BM, Ludke RL, Wakefield DS *et al.* Nonacute days of care within department of Veterans Affairs Medical Centers. *Med Care* 1991; **29 (suppl.):** AS51-AS63.
28. Restuccia JD, Holloway DC. Barriers to appropriate utilization of an acute facility. *Med Care* 1976; **14:** 599–573.
29. DeCoster C, Roos NP, Carriere KC, Peterson S. Inappropriate hospital use by patients receiving care for medical conditions: targeting utilization review. *Can Med Assoc J* 1997; **157:** 889–896.
30. Eggli Y, Yalcin T, Basterrechea L. *Executive Information System of the Hospices Cantonaux* Raisons de Santé, No. 5 (in French). Lausanne: Hospices cantonaux, 1997.
31. Schluep M, Bogousslavky J, Regli F *et al.* Justification of hospital days and epidemiology of discharge delays in a department of neurology. *Neuroepidemiology* 1994; **13:** 40–49.
32. 3M Health Information Systems. *All Patients Diagnosis Related Groups (AP-DRGs), Version 12.0. Definitions Manual.* Wallingford, CT: 3M, 1994.
33. Tilquin C. *PRN 87: Measure of Patients' Required Nursing Care in a 24-hour period* (in French). Montréal: EROS, 1987.
34. Jenicek M. Decision analysis and decision making in medicine, beyond intuition, guts and flair. In Eggli Y, Seker E, Wasser-Fallew, JB, eds, *Epidemiology the Logic of Modern Medicine.* Montréal: EPIMED International, 1995: pp. 299–325.
35. Guillain H, Brioschi Levi H, Eggli Y. *Bed Capacity of the Centre Hospitalier Universitaire Vaudois in 1998* (in French). Lausanne: Hospices cantonaux, in press.
36. Sox HC, Blatt MA, Higgins MC, Marton KI. Measuring the accuracy of clinical data. In Sox HC, ed., *Medical Decision Making.* Stoneham, MA: Butterworth publishers, 1988: pp. 103–145.
37. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd edn. New York: John Wiley & Sons, 1981.
38. Miettinen OS. Element of data analysis and inference. In *Theoretical Epidemiology: Principles of Occurrence Research in Medicine.* Albany, NY: Delmar publishers Inc, 1985: pp. 107–128.
39. Sox HC, Blatt MA, Higgins MC, Marton KI. Understanding new information: Bayes' theorem. In Sox HC, ed., *Medical Decision Making.* Stoneham, MA: Butterworth publishers, 1988: pp. 67–101.
40. Booth BM, Ludke RL, Wakefield DS *et al.* Non acute inpatient admissions to Department of Veterans Affairs Medical Centers. *Med Care* 1991; **29:** AS40–AS50.
41. Le Gall JR, Loirat P, Alperovitch A *et al.* A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; **12:** 975–977.
42. Robbins JA, Donaldson LJ. Analysing stages of care in hospital stay for fractured neck of femur. *Lancet* 1984; **2:** 1028–1029.
43. Smith CB, Goldman RL, Martin DC *et al.* Overutilization of acute care beds in Veterans Affairs hospitals. *Med Care* 1996; **34:** 85–96.
44. Mozes B, Halkin H, Katz A *et al.* Reduction of redundant hospital stay through controlled intervention. *Lancet* 1987; **1:** 968–969.
45. Wickizer TM. The effect of utilization review on hospital use and expenditures: a review of the literature and an update on recent findings. *Med Care Rev* 1990; **47:** 327–363.
46. Paranjpe N, Strumwasser I, Ronis DL *et al.* Efficiency gains in utilization review. *Qual Assur Util Rev* 1989; **4:** 108–114.
47. Strumwasser I, Paranjpe N, Ronis DL *et al.* Focused utilization review strategies. *Qual Rev Bull* 1989; **15:** 246–254.
48. Siu AL, Sonnenberg FA, Manning WG *et al.* Inappropriate use

of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986; **315:** 1259–1266.

49. Siu AL, Manning WG, Benjamin B. *Patient, Provider and Hospital Characteristics Associated with Inappropriate Hospitalization.* Santa Monica, CA: The RAND Corporation, 1990.

50. Blanc T. *Predicting inappropriate hospital use for focused concurrent utilization review.* Master's thesis. University of Washington: Seattle, 1993.

51. McMahon LF Jr, Newbold R. Variation in resource use within Diagnosis-Related Groups: the effect of severity of illness and physician practice. *Med Care* 1986; **24:** 388–397.

52. Donabedian A, ed. *Explorations in Quality Assessment and Monitoring: the Criteria and Standards of Quality*, Vol 2. Ann Harbor, MI: Health Administration Press, 1982.

Appendix A

The Appropriateness Evaluation Protocol (AEP) says that a hospital day is appropriate if any one criterion on the following list is fulfilled. The criteria are divided into three types of factor: medical services (criteria 1–11), nursing services (criteria 12–19) and patient characteristics (criteria 20–39). Most criteria justify the day of occurrence, but some apply also to the following day (criteria 21–23, 29,30,33–36) and others during the following 14 days (criteria 20, 28).

1 Procedure in operating room that day
2 Scheduled for procedure in operating room the next day, requiring preoperative consultation or evaluation
3 Cardiac catheterization that day
4 Angiography that day
5 Biopsy of an internal organ that day
6 Thoracocentesis or paracentesis that day
7 Invasive central nervous system diagnostic procedure that day
8 Any test requiring strict dietary control, for the duration of the diet
9 New or experimental treatment requiring frequent adjustments of dosage under direct medical supervision
10 Close monitoring by a doctor at least three times daily
11 Postoperative day for any procedure covered in criteria 1 or 3–7
12 Continuous monitoring of vital signs every 30 minutes, for at least 4 hours
13 Close medical monitoring by a nurse at least three times daily, under a doctor's orders
14 Intake and output measurement
15 Respiratory care, intermittent or continuous respirator use or inhalation therapy (with chest physical therapy, intermittent positive pressure breathing) at least three times daily
16 Parenteral therapy: intermittent or continuous intravenous fluid with any supplementation (electrolytes, protein or medications)
17 Intramuscular or subcutaneous injections, or both, at least twice daily
18 Care of major surgical wound and drainage (chest tubes, T tubes, Hemovacs, Penrose drains)
19 Treatment in an intensive care unit
20 Documented new acute myocardial infarction
21 Electrocardiographic evidence of acute ischaemia
22 Ventricular fibrillation
23 Transfusion due to blood loss
24 Active bleeding
25 Pulse rate <50 or >140 per minute
26 Blood pressure: systolic <90 or >200 mm Hg or diastolic <60 or >120 mm Hg
27 Severe electrolyte or blood gas abnormality
28 Cerebrovascular accident
29 Coma: unresponsiveness for at least 1 hour
30 Acute confusional state not due to alcohol withdrawal
31 Sudden onset of unconsciousness or disorientation
32 Acute loss of sight or hearing
33 Progressive acute neurological difficulties
34 Acute or progressive sensory, motor, circulatory or respiratory embarrassment
35 Acute haematological disorders, neutropenia, anaemia, thrombocytopenia, leucocytosis, erythrocytosis or thrombocytosis, yielding signs or symptoms
36 Fever of at least 101°F rectally, if patient was admitted for a reason other than fever
37 Persistent fever with rectal temperature >101°F for more than 5 days
38 Wound dehiscence or evisceration
39 Inability to void or move bowels, not attributable to a neurological disorder

Appendix B

Override decision table

Care required or medical event criteria	Care/medical event criteria combinations[1]					
	I	II	III	IV	V	VI
One or more of the AEP criteria 1–7, 9–12,14,15,18–27,29–39[2]	1	0	0	0	0	0
Supplementary criteria[3] (difficult investigation or physiotherapy, necessary observation, humanitarian reasons)	n.r.[3]	1	0	0	0	0
Cerebrovascular accident[4] (AEP criterion 28, stroke < 11 days or transient cerebral ischaemia < 5 days)	n.r.	n.r.	1	0	0	0
Single criterion specific to nursing care[5] among AEP criteria 8,13,16,17[6]	n.r.	n.r.	n.r.	1	1	0
Services requiring inpatient nursing[7]	n.r.	n.r.	n.r.	1	0	n.r.
Decision: Is the day justified?	yes	yes	yes	yes	no	no

AEP = Appropriateness Evaluation Protocol.

[1] 1 = care need or medical event present, 0 = care need or medical event absent, n.r. = presence/absence of care need or medical event not relevant to override decision.

[2] See Appendix A.

[3] An examination or a physiotherapy treatment are considered difficult if not practicable in an ambulatory nursing care setting, because the patient's illness or home care are not conducive to these forms of care. An observation is necessary if precise clinical findings are indispensable for choosing the best treatment, particularly in a neurological context. Humanitarian criteria include patients whose last days would involve great suffering if they were not treated in hospital.

[4] The number of days justified by a stroke or a transient cerebral ischaemia (criteria 28) was limited to 10 (instead of 14) and 4 (instead of 14) respectively.

[5] Some AEP criteria (8,13,16,17) are considered insufficient on their own if nursing could have been carried out by ambulatory care nurses (see below).

[6] See Appendix A.

[7] For instance, the following nursing services, according to the PRN classification [33], were considered specific to inpatient care: ventriculostomy care, tracheostomy care, manual ventilation, crisis intervention, more than one drainage with aspiration during the day, more than two catheterizations per day, more than three i.v./i.m./s.c./i.d. medications per day, more than four respiratory investigations per day, more than six irrigations per day, more than 14 per os medications, at least 20 intake or output check-ups per day. The full list is available upon request from the first author.

Appendix C

Per day average marginal costs (US$ estimation)

Activities	Costs (US$)
Nursing	62.5
Medical follow-up	12.5
Laundry	6.5
Meals	18.5
Total[1]	100.0

Source: Financial office of the Centre Hospitalier Universitaire Vaudois, Switzerland.

[1] Operating room procedures, laboratory, radiology and physiotherapy costs are not included in the calculation, since they are generally linked to appropriate days.

The cost of the review is ~6.5 US$ per reviewed day. Considering that the proportion of avoidable days is 50% of inappropriate days, the cost–benefit ratio was $0.14 = 6.5/[0.5 \times (100–6.5)]$.

Accepted for publication 20 April 2001