

How valid are utilization review tools in assessing appropriate use of acute care beds?

Norman Kalant,* Marc Berlinguet,† Jean G. Diodati,**
Leonidas Dragatakis,†† François Marcotte††

Abstract

Background: Despite their widespread acceptance, utilization review tools, which were designed to assess the appropriateness of care in acute care hospitals, have not been well validated in Canada. The aim of this study was to assess the validity of 3 such tools — ISD (Intensity of service, Severity of illness, Discharge screens), AEP (Appropriateness Evaluation Protocol) and MCAP (Managed Care Appropriateness Protocol) — as determined by their agreement with the clinical judgement of a panel of experts.

Methods: The cases of 75 patients admitted to an acute cardiology service were reviewed retrospectively. The criteria of each utilization review tool were applied by trained reviewers to each day the patients spent in hospital. An abstract of each case prepared in a day-by-day format was evaluated independently by 3 cardiologists, using clinical judgement to decide the appropriateness of each day spent in hospital.

Results: The panel considered 92% of the admissions and 67% of the subsequent hospital days to be appropriate. The ISD underestimated the appropriateness rates of admission and subsequent days; the AEP and MCAP overestimated the appropriateness rate of subsequent days in hospital. The kappa statistic of overall agreement between tool and panel was 0.45 for ISD, 0.24 for MCAP and 0.25 for AEP, indicating poor to fair validity of the tools.

Interpretation: Published validation studies had average kappa values of 0.32–0.44 (i.e., poor to fair) for admission days and for subsequent days in hospital for the 3 tools. The tools have only a low level of validity when compared with a panel of experts, which raises serious doubts about their usefulness for utilization review.

Reducing the time that patients spend in hospital is seen as one way to control health care costs. One procedure to demonstrate inefficiency in the use of acute care hospitals is utilization review, using a tool designed to assess the appropriateness of hospital admission and subsequent days spent in hospital by the retrospective application of objective criteria. Three tools widely used for this procedure are the ISD criteria set (Intensity of service, Severity of illness, Discharge screens¹), the AEP (Appropriateness Evaluation Protocol²) and the MCAP (Managed Care Appropriateness Protocol³). The need for hospital admission is determined daily by the application of a set of explicit criteria; if these are met for a given day, then that day spent in an acute care hospital is considered appropriate. Although designed for managing care during time spent in hospital, the tools have also been used retrospectively to assess resource utilization.

The ISD consists of sets of diagnosis-independent criteria applicable to specified levels of care (e.g., critical, acute, subacute) and to different body systems. Each set consists of 3 sections: severity of illness, intensity of service and discharge screens. The AEP and the MCAP, which was derived from the AEP, each consist of a set of admission criteria and a set of day-of-care criteria related to patient severity of illness and the clinical services required. The criteria are independent of diagnosis or body system and are applicable to all patients.

These tools have been used for utilization review in most provinces. Butler and

Research

Recherche

From *the Department of Medicine and †the Division of Cardiology, Sir Mortimer B. Davis – Jewish General Hospital, Montreal, Que., and ††the Ministère de la Santé et des Services Sociaux du Québec, Quebec City, Que.

This article has been peer reviewed.

CMAJ 2000;162(13):1809-13

‡ See related article page 1824

associates⁴ studied the reliability and validity of the AEP. The MCAP was used in 2 institutional studies^{5,6} and the ISD was used in a number of provincial surveys⁷⁻¹³ to determine the rate of inappropriate use of acute care beds. In several cases the tool criteria were reviewed before the study to ensure that they reflected local practice, but in none of the studies with the ISD and the MCAP was the tool validated by comparison with decisions of appropriateness obtained by a “gold standard,” namely a panel of physicians. The purpose of our study was to assess the validity of the 3 tools as determined by the level of agreement with the judgement of a panel of experts.

We used the 1997 versions of the ISD and the MCAP and the last version of the AEP (AEP criteria are restated with no substantive changes, but with notes and clarifications of each criterion in an unpublished manual supplied by Dr. Restuccia, dated May 1998¹⁴).

Methods

Seventy-five consecutive patients admitted during March 1998 to the coronary care unit (CCU) with a provisional diagnosis of acute myocardial infarction or unstable angina were considered.

A health record analyst trained as a primary reviewer for the ISD, and with prior experience in its application to cardiac patients, assessed the admission day and each subsequent day spent in hospital according to the criteria for care at the level at which the patient was being treated (CCU, telemetry unit or acute cardiac unit). A senior nurse with extensive experience in the cardiac unit was trained as a primary reviewer in the use of the MCAP and determined the appropriateness of each day in hospital using that tool. Although the reviews were carried out during each patient's stay in hospital, there was no contact with the attending physicians and no influence on patient care. Daily reviews were ended when the patient was discharged or transferred from the acute care nursing unit. A physician reviewer verified the application of the criteria but did not overrule any decisions on the basis of other information or clinical judgement.

Several months later the primary reviewer of the MCAP reassessed the hospital stays using the AEP criteria. A physician who was unaware of the appropriateness ratings obtained with the tools prepared a summary of each patient's stay using a standardized, day-by-day format, and included information on the day on which it became available to the attending physician and residents as well as information contained in the progress notes from the patient's record (copied verbatim when feasible). The summaries were reviewed independently by 3 cardiologists, none of whom was involved in the care of these patients. The cardiologists used their clinical judgement to assess the appropriateness of the admission and of each day in hospital. They then met to discuss their assessments to reach a consensus. However, complete agreement was not essential, and in such cases the majority opinion was accepted as the panel's judgement.

Each tool was assessed by the level of overall agreement between the tool and the panel as determined by the kappa statistic¹⁵ and by the specific inappropriate agreement (the number of days for which care at an acute care level was judged inappropriate by *both* the tool and the panel divided by the number of days judged inappropriate by *one or both* the tool and the panel).¹⁶ Kappa has a value of zero if agreement is due to chance alone and a value of 1

if agreement is perfect. Thus, 0–0.4 indicates poor agreement, 0.4–0.75 indicates fair to good agreement and 0.75–1.0 indicates good to perfect agreement.¹⁷ Sensitivity and specificity were calculated for each tool, using the panel's judgement as the “gold standard,” with “inappropriate” as the equivalent of “positive.” Thus, sensitivity is the “true inappropriate” rate of a tool (the proportion of the days considered inappropriate by the panel that were also judged inappropriate by the tool), and specificity is its “true appropriate” rate. Other statistical procedures included the χ^2 test for differences between proportions and 2-way analysis of variance (unequal subgroup numbers).

Results

The number of days judged appropriate by the tools and the panel are shown in Table 1. Compared with the panel's findings, the number of appropriate days was underestimated by the ISD and overestimated by the MCAP and the AEP, for both admission days and subsequent days. The rates of appropriateness of hospital admission differed significantly between the ISD and the AEP and MCAP ($p < 0.02$), and the rates of appropriateness were lower for subsequent days in hospital than for admission days ($p < 0.01$). The proportion of days judged inappropriate increased with increasing length of stay for about 6 days and then became stable (Fig. 1), for both panel and tools. For 33 patients with a final diagnosis of myocardial infarction the panel and each tool essentially judged all admissions to be appropriate; the assessments of the appropriateness of subsequent days were virtually identical to those for the whole group.

The level of agreement between the tools and the panel, as shown by the kappa values, was low, except for the ISD in its assessment of admission days (Table 2). The agree-

Table 1: Number of days in acute care hospitals judged appropriate by utilization review tools and a panel of physician experts

Tool/panel	No. (and %) of days deemed appropriate	
	Admission	Subsequent stay in hospital
All patients (n = 75)		
ISD	60 (80)*	184 (40)†
MCAP	75 (100)	370 (81)†
AEP	75 (100)	390 (85)†
Panel	69 (92)	309 (67)
Patients with myocardial infarction (n = 33)		
ISD	32 (97)	96 (41)†
MCAP	33 (100)	185 (79)‡
AEP	33 (100)	198 (85)†
Panel	33 (100)	154 (66)

Note: ISD = Intensity of service, Severity of illness, Discharge screens, MCAP = Managed Care Appropriateness Protocol, AEP = Appropriateness Evaluation Protocol.

* $p < 0.05$, † $p < 0.001$, ‡ $p < 0.005$, for difference between tool and panel.

ment between the MCAP and the AEP was good; this result was expected because the MCAP was derived from the AEP. With the panel's judgements taken to be "correct," all tools showed a sensitivity (true inappropriate rate) of 1.00 and a specificity (true appropriate rate) of 0.87 or higher for admission days. For subsequent days the ISD showed a sensitivity of 0.94 and a specificity of 0.57; the corresponding values were 0.33 and 0.87 for the MCAP and 0.29 and 0.92 for the AEP. Before the consensus meeting, the kappa value for agreement between individual physicians and each of the tools was 0.11 on average, and between pairs of physicians 0.32 on average.

Interpretation

In previous Canadian studies⁴⁻¹³ an average of 39% (range 8%–73%) of admissions and 49% (range 25%–82%) of hospital days were judged inappropriate by one of the utilization review tools. Although our inappropriateness rates were well within these ranges, comparison is of questionable value because of the patient populations studied. However, one report¹⁰ included a subgroup of patients with myocardial infarction, assessed by the ISD: 96% of admissions but only 36% of subsequent days were deemed appropriate. Our results for such patients were very similar (Table 1).

The importance of these rates depends on the validity of the utilization review tools, which are assessed by comparing the "decisions" of the tools with the judgement of a panel of experts. We searched the MEDLINE and HEALTHSTAR databases for independent validation studies providing kappa estimates of the agreement between tool and panel judgements for adult patients. Because the criteria have been modified over the years, the search was limited to the past 10 years. All of the reports found were based on the ratings of a primary reviewer. For both the ISD and the AEP and MCAP (considered together) the kappa values for overall agreement with the expert panels varied considerably but on average showed only

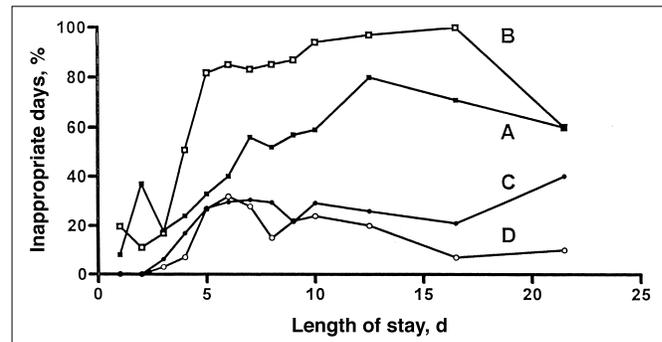


Fig. 1: Relation between length of stay in hospital and proportion of days found to be inappropriate by (A) a panel of physician experts and 3 utilization review tools: (B) Intensity of service, Severity of illness, Discharge screens (ISD), (C) Managed Care Appropriateness Protocol (MCAP) and (D) Appropriateness Evaluation Protocol (AEP). Because of the small number of patients remaining in hospital for more than 10 days, data were pooled for days 11–14, 15–18 and 19–24.

Table 2: Level of agreement between expert panel and utilization review tools

Variable	ISD		MCAP		AEP	
	κ value	SIA, %	κ value	SIA, %	κ value	SIA, %
Admission days						
Panel	0.52	40	*	*	*	*
Subsequent days						
Panel	0.42	50	0.22	25	0.24	24
ISD	–	–	0.23	29	0.20	24
MCAP	–	–	–	–	0.70	60
All days						
Panel	0.45	49	0.24	25	0.25	23
ISD	–	–	0.24	28	0.21	23
MCAP	–	–	–	–	0.71	60

Note: SIA = specific inappropriate agreement (see Methods for details).
*All other values in this matrix were 0.

Table 3: Reported levels of appropriateness of days in hospital and of validity of utilization review tools*

Tool†	Admission days			Subsequent days		
	Mean level of appropriateness (and SEM), %	κ value	SIA	Mean level of appropriateness (and SEM), %	κ value	SIA
ISD	63 (5) (n = 12)	0.41 (0.04) (n = 8)	48 (4) (n = 4)	47 (6) (n = 8)	0.44 (0.07) (n = 7)	60 (n = 3)
AEP + MCAP‡	76 (5) (n = 9)	0.32 (0.09) (n = 8)	37 (n = 2)	61 (5) (n = 10)	0.43 (0.07) (n = 8)	44 (12) (n = 4)

Note: SEM = standard error of the mean.

*For each tool the upper row of numbers = mean (and SEM) of reported values, including those of the present study; the lower row = number of reports. For fewer than 3 reports only the mean is given.

†References for the ISD tool = 5,7–13,16,18–22; references for the AEP and MCAP tools = 4–6,16, 22–32.

‡There were few reports on MCAP. Because the results were similar to those of AEP, the 2 sets of values were combined.

poor to fair validity (Table 3); specific inappropriate agreement was similar. Initially these tools were designed for managing care and included review by a physician of days rated inappropriate on primary review, with the option of overriding such ratings. Because all of the validation studies were based on the primary review alone, the results support the conclusion of Strumwasser and coworkers¹⁶ that the tools used in this way have too low a level of validity to justify their use as the sole basis for deciding the appropriateness of days in hospital. Because the past⁵⁻¹³ and current use of the tools for utilization review is also based on primary review alone, the estimated rates of inappropriate use of acute care hospitals should be viewed with some degree of scepticism.

There are 2 potential sources of bias in the results of our study. First, preparation of the patient abstracts involved selection of information that may have affected the panel's decision. Although we do not believe that any important data were omitted, the problem, common to all retrospective utilization review studies, remains. Second, the time lag before the application of the AEP tool may have introduced a systematic bias in the reviewer; we believe that this bias is negligible because a second reviewer monitored the application of the criteria.

Incorporation of a secondary review into the utilization review process might raise the level of agreement with the expert panel and so make the tool more acceptable. However, it was the frequent divergence of clinical opinion among individual physicians that led to the development of the tools: the goal was to replace the subjective use of implicit criteria by a physician reviewer with the objective application of explicit criteria of the tool.³³ Reported kappa values (including our results) for agreement between individual physician reviewers are on average 0.29 (standard error 0.06).^{16,20,22,34} This indicates that the addition of a secondary review by a single physician would reintroduce the source of variability that the tool was intended to eliminate and would add a major source of inconsistency in ratings. In contrast, agreement between *panels* of physicians³⁵⁻³⁷ is much stronger than between individual physicians, so secondary review of inappropriate days by an expert panel would increase the validity of the tool. However, the tools are relatively inefficient for assessing subsequent days in hospital, because of low sensitivity or low specificity. A focused review process³⁸ by a panel of physicians using a minimal representative patient sample³⁹ may be more efficient.

An additional concern is the lack of information about the predictive validity of the tools. The concept of an inappropriate day at an acute care level suggests that the quality of care would not be reduced at a subacute level. The accuracy of this assumption must be established by showing that patients discharged to a lower level of care, on the basis of the tool's criteria, have outcomes at least as good as those of patients "managed" by the attending physician using clinical judgement. There have been no reports of such trials.

Although utilization review tools are widely accepted these considerations taken together raise serious questions about the value of the tools as they are currently used and whether they should be used at all.

We thank H el ene Jean-Baptiste for acting as primary reviewer for the ISD tool, for record keeping and for tabulating results. We also thank Jennifer Eastmond for serving as primary reviewer for the MCAP and AEP tools and Fr ed eric Abergel for computer program development.

This project was supported by a grant from the Minist ere de la Sant e et des Services Sociaux du Qu ebec.

Competing interests: None declared.

References

1. *The InterQual review system*. Marlborough (MA): InterQual Products Group; 1996.
2. Gertman PM, Restuccia JD. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Med Care* 1981;19:855-71.
3. *The Managed Care Appropriateness Protocol*. Wellesley (MA): Oak Group; 1996.
4. Butler JS, Barrett BJ, Kent G, McDonald J, Haire R, Parfrey PS. Detection and classification of inappropriate hospital stay. *Clin Invest Med* 1996;19:251-8.
5. Stevens P. *Utilization review tools – an evaluation*. Vancouver: Simon Fraser Health, Simon Fraser University; 1998.
6. Kaplow M, Charest S, Mayo N, Benaroya S. Managing patient length of stay better using an appropriateness tool. *Health Manage Forum* 1998;11:13-6.
7. *Barriers to community care*. Saskatoon (SK): Health Services Utilization and Research Commission; 1994. Final report.
8. Sims C. *ISD project, utilization management program, CPU medicine*. Vancouver: Vancouver Hospital and Health Sciences Centre; 1995.
9. Cardiff K, Sheps SB, Thompson DM. To be or not to be in hospital? Vancouver: Centre for Health Services and Policy Research, University of British Columbia; 1996. Report no HPRU 96:11D.
10. Utilization Steering Committee, Joint Policy and Planning Committee. *Non-acute hospitalization project – final report*. Toronto: Ontario Ministry of Health; 1997. Reference document RD6-3.
11. DeCoster C, Roos NP, Carri ere KC, Peterson S. Inappropriate hospital use by patients receiving care for medical conditions: targeting utilization review. *CMAJ* 1997;157:889-96. Available: www.cma.ca/cmaj/vol-157/issue-7/0889.htm
12. Wright CJ, Cardiff K, Kilshaw M. *Acute medical beds: How are they used in British Columbia?* Vancouver: Centre for Health Services and Policy Research, University of British Columbia; 1997. Report no HPRU 97:7D.
13. Wright CJ, Cardiff K. *The utilization of acute care medical beds in Prince Edward Island*. Vancouver: Centre for Health Services and Policy Research, University of British Columbia; 1998. Report no HPRU 98:14D.
14. Restuccia JD, Payne SMC, Lenhart G, Constantine HP, Fulton JP. Assessing the appropriateness of hospital utilization to improve efficiency and competitive position. *Health Care Manage Rev* 1987;12:17-27.
15. Armitage P, Berry G. *Statistical methods in medical research*. Oxford (UK): Blackwell Scientific Publications; 1994. p. 443-5.
16. Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. *Med Care* 1990;28:95-109.
17. Fleiss J. *Statistical methods for rates and proportions*. 2nd ed. New York: Wiley & Sons; 1981. p. 212-36.
18. Ludke RL, MacDowell NM, Booth BM, Hunter SA. Appropriateness of admissions and discharges among readmitted patients. *Health Serv Res* 1990;25:501-25.
19. Coast J, Inglis A, Morgan K, Gray S, Kammerling M, Frankel S. The hospital admissions study in England: Are there alternatives to emergency hospital admission? *J Epidemiol Community Health* 1995;49:194-9.
20. Inglis AL, Coast J, Gray SF, Peters TJ, Frankel SJ. Appropriateness of hospital utilization. *Med Care* 1995;9:952-7.
21. Smith CB, Goldman RL, Martin DC, Williamson J, Weir C, Beauchamp C, et al. Overutilization of acute-care beds in Veterans Affairs hospitals. *Med Care* 1996;34:85-96.
22. Goldman RL, Weir CR, Turner CW, Smith CB. Validity of utilization management criteria for psychiatry. *Am J Psychiatry* 1997;154:349-54.
23. Winickoff RN, Restuccia JD, Fincke BJ. Appropriateness Evaluation Study Group. Concurrent application of the Appropriateness Evaluation Protocol to acute admissions in Department of Veterans Affairs medical centers. *Med Care* 1991;29(8 Suppl):AS64-76.

24. Booth BM, Ludke RL, Wakefield DS, Kern DC, du Mond CE. Relationship between inappropriate admissions and days of care: implications for utilization management. *Hosp Health Serv Adm* 1991;36:421-37.
25. Apolone G, Alfieri V, Braga A, Caimi V, Cestari C, Crespi V, et al. A survey of the necessity of the hospitalization day in an Italian teaching hospital. *Qual Assur Health Care* 1991;3:1-9.
26. Tsang P, Severs MP. A study of appropriateness of acute geriatric admissions and an assessment of the Appropriateness Evaluation Protocol. *J R Coll Physicians Lond* 1995;29:311-4.
27. Peiro S, Meneu R, Rosello ML, Portella E, Carbonell-Sanchis R, Fernandez C, et al. Validity of the protocol for evaluating the inappropriate use of hospitalization. *Med Clin (Barc)* 1996;107:124-9.
28. 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;8:543-53.
29. Smith HE, Pryce A, Carlisle L, Jones JM, Scarpello J, Pantin C. Appropriateness of acute medical admissions and length of stay. *J R Coll Physicians Lond* 1997;31:527-32.
30. Chopard P, Perneger TV, Gaspoz JM, Lovis C, Gousset D, Rouillard C, et al. Predictors of inappropriate hospital days in a department of internal medicine. *Int J Epidemiol* 1998;27:512-9.
31. Merom D, Shohat T, Harari G, Oren M, Green MS. Factors associated with inappropriate hospitalization days in internal medicine wards in Israel: a cross-national survey. *Int J Qual Health Care* 1998;10:155-62.
32. Doyle M, Barrett BJ, McDonald J, McGrath J, Parfrey PS. The efficiency of acute care bed utilization in Newfoundland and Labrador. *Health Manage Forum* 1998;11(2):15-9.
33. Restuccia JD. The evolution of hospital utilization review methods in the United States. *Int J Qual Health Care* 1995;7:253-60.
34. Wassertheil-Smoller S, Tobin J, Steingart R. Assessing the appropriateness of medical care [letter]. *N Engl J Med* 1999;339:1478.
35. Pearson SD, Margolis CZ, Davis S, Schreier LK, Sokol HN, Gottlieb LK. Is consensus reproducible? A study of an algorithmic guidelines development process. *Med Care* 1995;33:643-60.
36. McDonnell J, Meijler J, Kahan JP, Bernstein SJ, Rigger H. Panellist consistency in the assessment of medical appropriateness. *Health Policy* 1996;37:139-52.
37. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;338:1888-95.
38. Nelson ME, Christenson RH. The focused review process: a utilization management firm's experience with length of stay guidelines. *J Qual Improve* 1995;21:477-87.
39. Ash A, Schwartz M, Payne SMC, Restuccia JD. The self-adapting focused review system. Probability sampling of medical records to monitor utilization and quality of care. *Med Care* 1990;28:1025-39.

Reprint requests to: Dr. Norman Kalant, Department of Medicine, Sir Mortimer B. Davis – Jewish General Hospital, 3755 Chemin de la Côte-Sainte-Catherine, Montreal QC H3T 1E2; fax 514 340 7508; nkalant@mtd.jgh.mcgill.ca



Canadian Medical Association

133rd Annual Meeting

Saskatoon, Saskatchewan
13-16 AUGUST 2000

Hosted by the
Saskatchewan Medical Association

Registration and travel information:
CMA Meetings and Travel

800 663-7336 or
613 731-8610 x2383
Fax: 613 731-8047





ASSOCIATION
MÉDICALE
CANADIENNE

CANADIAN
MEDICAL
ASSOCIATION