

Validation of the paediatric appropriateness evaluation protocol in British practice

Ursula Werneke, Helen Smith, Iain J Smith, Jeanette Taylor, Roderick MacFaul

Abstract

The reliability and validity of the North American paediatric appropriateness evaluation protocol (PAEP) for use in paediatric practice in Britain was tested. The protocol was applied to 418 case records of consecutive emergency admissions to three Yorkshire district general hospitals. The PAEP ratings were then compared with a clinical consensus opinion obtained from two expert panels. Altogether 32% of the admissions were rated inappropriate by the PAEP and 36% by the panels. Validity of the PAEP, as measured by agreement beyond chance with the expert panel rating, was only moderate with a κ of 0.29 (95% confidence interval 0.11 to 0.47). The PAEP has limited validity for evaluating British paediatric practice. Utilisation review instruments developed in differing clinical cultures should be used with caution until shown to be valid for the practice setting under review.

(*Arch Dis Child* 1997;77:294-298)

Keywords: paediatric appropriateness evaluation protocol; hospitalisation; paediatric admissions

Evaluation of appropriateness of an acute hospital admission with measurement tools based on standardised criteria is a relatively new concern in the UK. In North America, such utilisation review instruments have been developed and used since the 1970s, aiming to limit costs and increase efficiency of acute care. The instruments use so called explicit, objective criteria developed for this purpose. One of the widely applied tools is the appropriateness evaluation protocol for use in adult practice (AEP).¹ The AEP criteria are based on the level of service provided and, to some extent, factors which identify severity of illness, independent of diagnoses. The protocols were designed for retrospective or concurrent application to case notes. Fulfilment of one of the relevant criteria would make that day of hospital care appropriate. The original criteria in the AEP were tailored to adult specialties. Two versions for children, both called the paediatric appropriateness evaluation protocol (PAEP) were later derived from the adult tool.^{2,3} The PAEP has been used in several countries, although the methods of application and sampling strategies have been inconsistent making results difficult to compare.⁴

Utilisation reviews are not designed as arbiters of individual patient care. However, results

of reviews can influence patient care, service provision, and health policy. If they are unreliable or invalid their use may adversely affect the quality of health care and unfairly penalise patients; for example admission may be discouraged though needed, or inpatient facilities erroneously reduced.⁵ On the other hand, evaluation of appropriateness of hospitalisation is a valuable method for monitoring services and informing planning or reconfiguration of services. It is essential, therefore, that before an audit tool is used, reliability and validity of the instruments are sufficiently high. Reliability is a measure of reproducibility. Interobserver reliability measures the extent to which raters independently arrive at the same results. Validity refers to the extent to which the instrument measures what it purports to measure. Validity of utilisation review instruments is dependent on the clinical culture in which an instrument is applied.¹ A review instrument should also be acceptable and credible to the clinicians whose practice is being evaluated so as to engage their cooperation when managing change. This paper reports a study of the reliability and validity of the PAEP in UK paediatric practice. It was conducted during a British Paediatric Association* study of paediatric admissions to district general hospitals in 1993/4 (*now the Royal College of Paediatrics and Child Health).

Method

The study population consisted of 426 consecutive paediatric admissions to three Yorkshire hospitals over two sample periods, one each of three weeks in summer and winter. The PAEP as developed by Kreger and Restuccia (see Appendix),³ was applied retrospectively by a research assistant (JT) to case notes of 418 of these children. Eight children were excluded because the case notes were not available.

JT was trained in the application of the PAEP by HS who had reported the application and validation of the instrument in a Canadian tertiary care hospital.⁶

RELIABILITY

An interobserver reliability study was carried out before and after the application. Initially, JT and HS each applied the PAEP to 50 sets of paediatric case notes randomly selected in a hospital in Southampton. The over-ride option was explained but not used, either in training or in application. After completion of the study, a second reliability check was conducted on 50 randomly selected case notes of paediatric admissions in Wakefield (table 1).

London Health
Economics
Consortium, London
School of Hygiene and
Tropical Medicine
U Werneke

Wessex Primary Care
Research Network,
Southampton
H Smith

Nuffield Institute for
Health, Leeds
I J Smith

Royal College of
Paediatrics and Child
Health, London
J Taylor
R MacFaul

Correspondence to:
Dr R MacFaul, Health
Services Committee, Royal
College of Paediatrics and
Child Health, 5 St Andrews
Place, Regents Park, London
NW1 4 LB.

Accepted 21 May 1997

Table 1 Interobserver reliability of the PAEP

	Before study (50 sets)	After study (50 sets)
Overall agreement (%)	92	94
Estimate of inappropriate admissions (JT) (%)	52	30
Estimate of inappropriate admissions (HS) (%)	48	32
Specific inappropriate agreement (%)	86	82
κ	0.72	0.86

Table 2 Clinical problems in the selected cases; values are %

Discharge diagnosis	Study overall (n=418)	Sample for PAEP validation (n=50)
Upper respiratory tract infection	15.2	16
Other	12.4	10
Gastroenteritis	10.1	18
Asthma	8.3	4
Febrile convulsion	8.1	10
Ingestion	6.8	4
Bronchiolitis	5.9	8
Viral infection	5.5	2
Croup	5.3	2
Convulsion	5.1	4
Gastrointestinal disorder	5.1	8
Lower respiratory tract infection	3.3	2
Specific infection (for example pertussis, varicella)	2.7	6
Abdominal pain	2.5	2
Urinary tract infection	1.4	0
Meningitis/septicaemia	1.0	0
Surgical	0.8	4
Soft tissue disorder	0.7	0

VALIDATION

The PAEP was subjected to a validation exercise based on the procedure used by Smith *et al.*⁶ Two panels of three paediatricians reviewed 25 sets of case notes each, thereby assessing a total of 50 case notes. Each panel was made up of three consultant paediatricians from other health regions not involved in the care of the patients.

SAMPLE

Fifty sets of notes was regarded as a sufficient sample for validation. This was based on two considerations: (1) the need to review a mixture of appropriate and inappropriate admissions. Given an estimated 25% prevalence of inappropriate admissions,⁴ a random sample of 45 cases was needed with an error margin of $\pm 12\%$, at a confidence level of 95% and (2) the validation exercise carried out in Canada limited the sample size to 50 cases in view of limited clinician time to participate in the exercise.⁶

STATISTICAL METHOD

The following statistical measures of reliability and validity were calculated⁵:

Table 3 Agreement between members of panel 1

	A v B	C v A	C v B
Overall (%)	68	80	72
Inappropriate rater 1 (%)	56	56	20
Inappropriate rater 2 (%)	32	36	32
Specific inappropriate agreement (%)	43	64	70
κ (95% confidence interval)	0.39 (0.19 to 0.60)	0.61 (0.43 to 0.79)	0.28 (0.08 to 0.49)

Table 4 Agreement of panel 1 with the PAEP

Overall (%)	64
PAEP inappropriate (%)	40
Panel inappropriate (%)	44
Specific inappropriate agreement (%)	40
κ (95% confidence interval)	0.26 (0.09 to 0.48)

Overall agreement: The proportion of judgments in which two reviewers or assessment methods (PAEP rater and expert panel) agree.

Specific inappropriate agreement: The proportion of admissions classified as inappropriate by both reviewers/assessment methods of the total admissions which were judged as inappropriate by at least one reviewer.

Cohen's κ : A correlation coefficient measuring agreement beyond chance.

CASE NOTE SELECTION

Case notes were initially stratified into PAEP appropriate and PAEP inappropriate according to the proportion (32.7% inappropriate) observed in the study. Then case notes were selected randomly from these two groups.

PANEL ASSESSMENTS

As the PAEP had been applied only to the day of admission, photocopies were made of notes relating only to that day. Before each of the two panels was assembled, each member individually reviewed 25 sets of anonymised notes and recorded a decision on whether the admission was needed in response to the following brief:

Based on the admission day and taking account of all the information available in the record and on services currently present, consider was the admission justified or needed? Please record your judgment in response to the question: did the patient require the services of an acute care setting on the day in question?

For each case, the majority view of each panel (2/3 or 3/3) was taken as the result. This was then compared with the rating of the PAEP for each case. In the next step, each panel was assembled and the cases were discussed by them. This provided an opportunity to understand why differences between the PAEP and expert panel judgments had arisen, and whether the PAEP criteria had been relevant to clinical paediatric practice as derived from the case notes.

Results

The clinical problems found in the selected cases were representative of the casemix admitted to the hospitals over the study periods (table 2) (M Stewart *et al.*, manuscript in preparation).⁷

RELIABILITY

Both studies looking at interobserver reliability yielded good to very good agreement with a κ of 0.72 and 0.86 respectively.

VALIDITY

The results are presented for the two panels first separately, and then as a summary.

(1) Panel 1

There were three raters A, B, and C. The agreement between each pair of panel mem-

Table 5 Agreement between members of panel 2

	D v E	F v E	F v D
Overall (%)	92	80	80
Inappropriate rater 1 (%)	28	32	32
Inappropriate rater 2 (%)	36	36	28
Specific inappropriate agreement (%)	77	55	50
κ (95% confidence interval)	0.81 (0.69 to 0.94)	0.55 (0.33 to 0.74)	0.52 (0.34 to 0.71)

Table 6 Agreement of panel 2 with the PAEP

Overall agreement (%)	68
PAEP inappropriate (%)	24
Panel inappropriate (%)	32
Specific inappropriate agreement (%)	27
κ (95% confidence interval)	0.21 (0.00 to 0.43)

bers using subjective (implicit) criteria alone was only fair. However, application of explicit criteria in the PAEP did not score higher since agreement based on the panel majority decision with the PAEP only yielded a κ of 0.26 (tables 3 and 4).

(2) Panel 2

There were three raters D, E, and F. Again, the agreement between each pair of panel members based only on implicit criteria (table 5) was higher than their overall agreement on majority decision based on explicit criteria from the PAEP (table 6).

SUMMARY OF RESULT OF VALIDATION EXERCISE

Each panel was assembled to discuss the cases. The first panel considered all nine cases in which the panel and the PAEP rating had been discordant (PAEP appropriate and panel inappropriate, five cases; PAEP inappropriate and panel appropriate, four cases). One change in decision followed this discussion. The second panel considered all eight cases where there was disagreement (PAEP appropriate and panel inappropriate, five cases; PAEP inappropriate and panel appropriate, three cases). No change in decision followed this discussion.

The final validity score was determined combining the results of the two panels were combined (table 7). The effect of the panel change in opinion on one case in panel 1 from 'inappropriate' to 'appropriate' was to alter the final κ score from 0.26 to 0.29. This is only a fair level of agreement.

CLINICAL ASPECTS OF PANEL DISCUSSIONS

The one case in which the panel changed their decision from inappropriate to appropriate in accordance with the PAEP result was a child with an upper respiratory infection and an immune defect. There were seven cases which the PAEP rated inappropriate but for which the panel judged admission needed. Significant

clinical problems were present in six. In the remaining child, parental anxiety was thought sufficiently high to justify admission. There were nine other cases in which the PAEP had rated the admission appropriate contrary to the panel ratings. In eight of these cases, the panel took the view that the patient should have been managed as an outpatient. One child with cerebral palsy who had a fit should have been treated in a respite care facility. Greater detail is given in table 8.

Discussion

Although interobserver reliability for the PAEP achieved good to very good results, only a fair level of validity (κ 0.29) was achieved judging the PAEP ratings against an expert panel as a gold standard.

In previous studies in other countries, the PAEP has been applied to mixtures of paediatric medical and surgical cases, emergency and elective admissions and to mixtures of secondary and tertiary care.⁴ The only other study which included a validation exercise based on expert panels related to children admitted to a tertiary care centre.⁶ A casemix of generally more severely ill children in whom appropriateness of admission is easier to rate may explain the higher κ of 0.68 observed in the Canadian study for day of admission. One factor that may have contributed to the higher levels of agreement between clinicians and the PAEP in that study was that they were asked to make a judgment taking account of hypothetical alternative arrangements which might have prevented the admission.

Kemper *et al* developed and applied their version of the PAEP independently.² Kemper reported a validation of appropriateness of days of care using sensitivity and specificity comparing their PAEP ratings against clinical judgments, and reported good 'face validity' (that is the criteria looked plausible and relevant). They showed that the PAEP was highly sensitive, identifying the truly appropriate days of care in 93% of all cases. Specificity of 78% was less favourable, that is, 22% of days were falsely categorised as inappropriate.⁷ A validation exercise was not included in the development of the other version of the PAEP by Kreger and Restuccia,³ which is the version used in our study and the Canadian one (J D Restuccia, personal communication). Validity scores for the original adult version were κ = 0.4 for admission and κ = 0.7 for day of care, but it is doubtful whether these can be presumed to apply to the PAEP. The PAEP has been modified for use in England by Esmail taking account of UK paediatric practice.⁸ High interobserver levels of reliability have been found for this modified tool, but a validity exercise using separate expert panels has not been carried out after its application (A Esmail, personal communication).

In our study, the PAEP was applied only to the day of admission. Subsequent days of care were not analysed because the length of stay of the children was short with half discharged after one day or less, and only 21% staying for longer than two days. The length of stay profile

Table 7 Combined agreement of panels 1 and 2 with the PAEP

	Before case discussion	After case discussion
Overall agreement (%)	66	68
PAEP inappropriate (%)	32	32
Panel inappropriate (%)	38	36
Specific inappropriate agreement (%)	38	36
κ (95% confidence interval)	0.26(0.08 to 0.44)	0.29(0.11 to 0.47)

Table 8 Clinical details in patients where panel and PAEP rating differed

Age	Clinical details and panel comment	Discharge diagnosis	PAEP criterion
Panel appropriate but PAEP inappropriate			
7 months	Rash and fever needed observation. Mother age 17 years and anxious	Viral infection	
6 months	Febrile convulsion and postictal	Febrile convulsion	
5 years	Fever and drowsy	Upper respiratory infection	
13 months	Diarrhoea, temperature 39°C, dehydrated	Gastroenteritis	
5 years	Head injury, parents very anxious	Head injury	
9 weeks	Diarrhoea and vomiting, mild dehydration	Gastroenteritis	
13 months	First febrile convulsion	Febrile convulsion	
PAEP appropriate but panel inappropriate			
3 weeks	Feeding problem*	Feeding problem	Need for observation
19 months	Scabies rash, parents requested admission*	Scabies	Not responding to outpatient treatment
16 months	Mild bronchiolitis*	Bronchiolitis	Intermittent or continuous respirator use at least every eight hours
4 months	Vomiting and failure to thrive*, weight normal on admission	Vomiting	Need for special observation or close monitoring of behaviour including energy intake in cases of failure to thrive
14 years	Fit in a child with cerebral palsy and severe learning disability who refused medication. Respite care could have prevented admission	Fit	Need for special observation or close monitoring of behaviour including energy intake in cases of failure to thrive
2 years	Given large dose of paracetamol by error*	Ingestion	Observation for toxic reaction to medication
22 months	Fever and cough*	Otitis media	Condition not responding to outpatient management
20 months	Upper respiratory infection with five week history*	Upper respiratory infection	Condition not responding to outpatient management
4 years	Swelling in neck with tonsillitis*	Streptococcal tonsillitis	Condition not responding to outpatient management

* Panel commented that patient could have been managed as outpatient.

observed in this study is typical of acute general paediatric practice in the UK. In North America, where the PAEP was developed and used, length of stay is generally longer. Also the definition of a paediatric admission may differ, with only those children who stay for 24 hours or more being classified as an admission, while children cared for in short stay emergency facilities are classified as ambulatory or outpatients. The UK paediatric inpatient population includes such patients, and application to the entire casemix may impair validity or produce results not relevant to UK practice. To our knowledge, the PAEP is not applied to children in short stay facilities in the United States.

The 'true' prevalence of inappropriate admissions in UK remains unknown. In situations in which there is no objective measure for the factor under study, validity should be measured through general coefficients of concordance like the κ coefficient rather than expressed through sensitivity and specificity.^{9 10} Caution should be exercised in basing the decision of whether to use a tool exclusively on grounds of the correlation coefficients reported. High overall agreement can be accompanied by low κ scores when the expected prevalence of the factor under study is either very high or very low. Therefore decision to use a tool should not be made exclusively on correlation coefficients but also be judged in terms of plausibility, relevance, and acceptability of the criteria it contains. The use of consensus panels of clinicians is one way to do this. It can be regarded as the 'the next best thing' to a true gold standard.⁵ As a gold standard it has limitations since variation between clinicians' judgments is generally high. In our study, a substantial difference was observed between members of each of the two panels with panel 2 achieving higher levels of agreement. Although the selection of panel 1 may have unduly disadvantaged the PAEP, the reverse may be the case for panel 2. Strumwasser *et al*¹¹

and others¹² have demonstrated that the result of a validity study can depend on the composition of the expert panels. Validity scores for the AEP changed substantially when the panels of experts working in a fee for service environment were replaced by physicians belonging to a health maintenance organisation where reimbursement is based on capitation. Similar observations were made by Inglis *et al* when validating another utilisation review instrument, the American intensity severity discharge protocol (ISD) in the UK,¹³ when higher κ scores were obtained when using a general practitioner panel as a gold standard compared with a specialist panel

Utilisation review tools developed in one health system may not be transferable to another. The study assessing the ISD, reported it to be not valid for routine assessment of hospital utilisation within the National Health Service taking account of the limited scope for alternative care arrangements.¹³

The results of our study show that the PAEP in its present form has limited validity, and therefore cannot be recommended for assessment of UK paediatric practice in general hospitals. Before introduction for routine use in the UK, utilisation review instruments developed in other countries for either adult or paediatric practice should undergo a formal assessment of both reliability and validity.

Funding: Department of Health.

- Restuccia JD. The evolution of hospital utilization review methods in the United States. *Int J Qual Health Care* 1995;7:253-60.
- Kemper KJ. Medically inappropriate hospital use in a pediatric population. *N Engl J Med*. 1988;318:1033-7.
- Kreger BE, Restuccia JD. Assessing the need to hospitalize children: pediatric appropriateness evaluation protocol. *Pediatrics* 1989;84:242-7.
- Werneke U, MacFaul R. Evaluation of appropriateness of paediatric admission. *Arch Dis Child* 1996;74:268-74.
- Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell JJ. Reliability and validity of utilization review criteria. *Med Care* 1990;28:95-111.
- Smith HE, Sheps S, Matheson DS. Assessing the utilisation of inpatient facilities in a Canadian pediatric hospital. *Pediatrics* 1993;92:587-93.

- 7 Kemper KJ, Fink HD, McCarthy PL. The reliability and validity of the pediatric appropriateness evaluation protocol. *Quality Review Bulletin* 1989;15:77-80.
- 8 Esmail A. The appropriateness of paediatric admission. London: University of London, July 1996. [PhD thesis.]
- 9 Feinstein AR, Cicchetti DV. High agreement but low kappa: the problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
- 10 Thompson DW, Walter WD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1988;41:949-58.
- 11 Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single disciplinary panel. *Health Serv Res* 1995;30:577-91.
- 12 Scott EA, Black N. Appropriateness of cholecystectomy in the UK—a consensus panel approach. *Gut* 1991;32:1066-70.
- 13 Inglis AL, Coast J, Gray SF, Peters TJ, Frankel SJ. Appropriateness of hospital utilisation: the validity and reliability of the intensity severity discharge review system in United Kingdom acute hospital setting. *Med Care* 1995;33:952-7.

Appendix

THE PAEP CRITERIA USED FOR ADMISSION DAY

(1) Surgery or procedure scheduled necessitating:

- (A) General or regional anaesthesia
- (B) Use of equipment, facilities, or procedure available only in a hospital.
- (2) Treatment in an intensive care unit.
- (3) Vital sign monitoring every two hours or more often (may include telemetry or bedside cardiac monitor).

(4) Intravenous medications and/or fluid replacement (does not include tube feedings).

(5) Intramuscular observation for toxic reaction to medication.

(6) Intramuscular antibiotics at least every eight hours.

(7) Intermittent or continuous respirator use at least every eight hours.

(8) Severe electrolyte/acid-base abnormality (any of the following values):

- (A) Sodium <123 mmol/l or >156 mmol/l
- (B) Potassium <2.5 mmol/l or >5.6 mmol/l
- (C) Carbon dioxide combining power (unless chronically abnormal) <20 mmol/l or >36 mmol/l
- (D) Arterial pH <7.30 or >7.45.

(9) Acute loss of sight or hearing within 48 hours.

(10) Acute loss of ability to move body part within 48 hours.

(11) Persistent fever greater than 37.8°C orally or 38.3°C rectally for more than 10 days.

(12) Active bleeding.

(13) Wound dehiscence or evisceration.

(14) Pulse rate greater or less than the following ranges (optimally a sleeping pulse for <12 years old):

6 months–1 year 364 days, 80–200/min

2–6 years of age, 70–200/min

7–11 years of age, 60–180/min

>12 years of age, 50–140/min.

(15) Blood pressure values falling outside following ranges:

6 months–1 year 364 days, 70–120/40–85 mm Hg

2–6 years of age, 75–125/40–90 mm Hg

7–11 years of age, 80–130/45–90 mm Hg

>12 years of age, 90–200/60–120 mm Hg.

(16) Acute confusional state, coma or unresponsiveness.

(17) Packed cell volume less than 30.

(18) Need for lumbar puncture, where this procedure is not done routinely on an outpatient basis.

(19) Condition not responding to outpatient management (specify):

(A) Seizures

(B) Cardiac arrhythmia

(C) Bronchial asthma or croup

(D) Dehydration

(E) Encopresis (for clear out)

(F) Other physiological problem.

(20) Special paediatric problems:

(A) Child abuse

(B) Non-compliance with necessary therapeutic regimen

(C) Need for special observation or close monitoring of behaviour including energy intake in cases of failure to thrive.