

원저

# 적절성 평가지침과 이유목록의 적용 가능성 평가

신영수<sup>1)</sup>, 김용익<sup>1)</sup>, 김창엽<sup>1)2)</sup>, 김윤<sup>1)2)</sup>, 김은경<sup>2)</sup>, 송윤미<sup>3)</sup>, 이영성<sup>4)</sup>  
서울대학교 의과대학 의료관리학교실<sup>1)</sup>, 서울대학교병원 QA 연구반<sup>2)</sup>,  
단국대학교 의과대학 가정의학교실<sup>3)</sup>, 한국의료관리연구원<sup>4)</sup>

## Applicability of Appropriateness Evaluation Protocol and Delay Tool

Youngsoo Shin,<sup>1)</sup> Yong-ik Kim,<sup>1)</sup> Chang-Yup Kim,<sup>1)2)</sup> Yoon Kim,<sup>1)2)</sup> Eun Gyung Kim,<sup>2)</sup> Yun Mi Song,<sup>3)</sup> Young Seong Lee<sup>4)</sup>  
Department of Health Policy and Management, College of Medicine, Seoul National University<sup>1)</sup>,  
QA Office, Seoul National University Hospital<sup>2)</sup>  
Department of Family Medicine, College of Medicine, Dankook University<sup>3)</sup>  
Korea Institute of Health Services Management<sup>4)</sup>

**Abstract Background:** An appropriate use of hospital beds can improve productivity of hospital significantly. The authors' previous study revealed that approximately one third of Korean hospital bed days and one sixth of admissions were inappropriately used, when it was measured by Appropriateness Evaluation Protocol(AEP) and Delay Tool modified into Korean situation by the authors. This study aims to evaluate applicability of the instruments in a new hospital. More specifically the study aims to measure appropriateness of the instruments used by newly trained nurse

reviewers at a new hospital setting.

**Methods:** In order to evaluate applicability of these instruments, agreement rates of the scores recorded by newly trained nurse reviewers with by skilled nurse reviewer and also compared with the scores recorded by physician's implicit decision were assessed. Agreement rates were derived from concurrent application of AEP and Delay Tool to 52 admissions and 104 patient days from internal medicine, pediatrics, and general surgery of one university hospital. Overall agreement rate, specific nonacute agreement rate, and kappa statistics were used to indicate level of agreement.

**Results:** Overall agreement rates on appropriateness between newly trained nurse reviewers and skilled nurse reviewer were 100% in admission and 98% in bed days. Overall agreement rates on reason for inappropriateness between newly trained nurse reviewers and skilled nurse reviewer were 96% in admission and 91% in bed days. Overall agreement rates between newly trained nurse reviewers and physician reviewer were 86% in admission and 87% in bed days.

**Conclusion:** Results indicated that AEP and Delay Tool were applicable to a new hospital in detecting inappropriate utilization of beds and reasoning of the inappropriateness. These instruments could contribute to enhance efficiency of hospital use, through continuous monitoring of level of inappropriate hospital use at national or individual hospital level.

**Key Words:** applicability, Appropriateness Evaluation Protocol, Delay Tool, hospital use

## I. 서론

의학적으로 불필요한 의료이용이 줄어들면 국가적 차원에서는 제한된 의료자원을 효율적으로 사용할 수 있고, 개별 의료기관 수준에서는 경영 수지를 개선할 수 있다. 또한 이를 통하여 환자에게 추가적인 병원 비용의 발생을 막을 수 있을 뿐만 아니라, 병원 감염과 같은 의원성 질환(iatrogenic disease)의 발생 위험을 줄일 수 있다.<sup>1)</sup> 우리 나라에서 병원 입원의 약 1/6 재원일의 약 1/3 정도가 부적절한 의료

이용이라는 연구 결과<sup>2)</sup>는 이러한 부적절한 입원 및 재원을 줄이기 위한 노력이 매우 시급하고 중요함을 시사하고 있다.

부적절한 입원 및 재원을 줄이기 위해서는, 우선 이러한 부적절한 의료이용을 찾아내고 그 이유를 밝히는 데 유용한 의료이용검토 도구가 있어야 한다. 적절성 평가지침(Appropriateness Evaluation Protocol: AEP)과 이유목록(Delay Tool)은 신뢰도와 타당도가 높아 유용한 의료이용검토 도구로 알려져 있다.<sup>3)-5)</sup> 적절성 평가지침은 원래 미국에서 개발되었으나, 현

재 다른 나라들에서도 이를 적용하여 병상이용의 적절성을 평가하는 연구가 활발하게 이루어지고 있다.<sup>6)~14)</sup> 국내 일부 병원에서도 이들 도구를 적용하여 병상 이용의 적절성 수준을 정기적으로 모니터링 하려 하고 있다.

적절성 평가지침과 이유목록을 병상 이용의 적절성 수준을 모니터 하는 도구로 실용화하기 위해서는 도구 자체의 적용 가능성, 즉 개별 병원의 어떤 인력을 검토자로서 활용할 수 있는가가 평가되어야 한다. 왜냐하면 외국 경험을 고려할 때, 우리 나라에서도 이러한 의료이용검토의 활동에 의사인력 이외의 인력을 활용해야 할 것으로 예상되기 때문이다. 이들 도구를 실용화하는 데에는 도구 자체의 신뢰도와 타당도, 그리고 적용 가능성 이외에도 병원 내의 여러 환경 요인들과 보다 높은 차원에서는 의료제도 환경 요인들과 같은 도구 외적인 요인들이 관련되어 있다. 그러나 우선은 도구 자체의 적용 가능성을 검토할 필요가 있다. 왜냐하면 의료제도 환경, 병원 내 요인 등은 우선 도구의 적용 가능성이 검증된 이후의 검토 대상이며, 의사결정과 관련된 요인들로 그 영향을 평가하기가 매우 어렵기 때문이다.

국내에서도 적절성 평가지침과 이유목록을 적용한 연구가 몇차례 이루어진 바 있다.<sup>2), 15), 16)</sup> 그러나 이들 연구에서는 의사를 검토자로 활용하였기 때문에, 의사 이외의 인력을 검토자로 활용할 경우에 있어서 적절성에 대한 판단의 신뢰도 및 타당도의 문제에 대한 평가는 아직까지 이루어지지 못했다. 외국에서의 연구 결과에 의하면 적절성 평가지침을 적용한 의사 검토자와 간호사 검토자간의 일치도 수준은 의사 검토자간의 일치도 수준과 유사하게 높은 것으로 알려져 있다.<sup>17)</sup> 따라서 우리 나라에서도 의사 이외의 인력이 검토자로서 활용될 가능성은 충분하다고 생각된다.

본 연구는 개별 병원에서 현지 간호사 인력을 적절성 평가지침과 이유목록의 검토자로 활용할 수 있는가를 평가함으로써 이들 도구 자체의 적용 가능성을 검증하고자 하였다. 이를 위하여 먼저 현지 간호사 인력에 대한 단기간의 교육을 실시한 후에 이들

피교육자와 숙련검토자와의 입원 및 재원일 적절성과 부적절성의 이유에 대한 판단의 일치도를 평가하였다. 이와 함께 이들 피교육자가 적절성 평가지침을 적용하여 입원 및 재원일 적절성을 판단한 결과와 임상적 판단과의 일치도를 평가하였다.

## II. 연구 방법

### 1. 연구 체계

본 연구는 1개 대학병원을 대상으로 적절성 평가지침과 이유목록을 적용한 '병상 이용의 적절성 평가'를 시행하는 과정에서 이루어졌다. 이를 위하여 본 연구진은 조사 대상 병원에 근무하는 간호사 2인을 대상으로 적절성 평가지침과 이유목록에 대한 교육을 실시하였다. 교육 후 피교육자 간호사들은 조사 대상 병원에서 '병상 이용의 적절성 평가'를 시행하였으며, 이는 조사 기간 내에 새로이 입원한 환자를 퇴원할 때까지 추적 조사하는 방법으로 이루어졌다.

이러한 '병상 이용의 적절성 평가' 과정에서 피교육자 간호사와 숙련검토자 간호사간의 입원 및 재원일 적절성과 부적절성의 이유에 대한 판단의 일치도를 평가하였다. 피교육자 간호사들간의 판단의 일치도는 따로 평가하지 않았다. 이와 함께 피교육자 간호사가 적절성 평가지침을 적용하여 판단한 결과와 임상 의사가 묵시적 기준(implicit criteria)에 의거하여 판단한 결과간의 일치도를 평가하였다. 피교육자의 적절성 및 부적절성의 이유에 대한 판단 결과를 이러한 판단의 참값(gold standard)에 가까운 숙련검토자 간호사 또는 임상 의사에 의한 묵시적 판단과 비교함으로써 피교육자 간호사를 이들 도구의 검토자로서 활용할 수 있는가를 평가하고자 하였다.

본 연구의 체계는 Fig. 1과 같다.

### 2. 연구 대상

본 연구의 조사 대상 병원은 서울 지역에 위치한

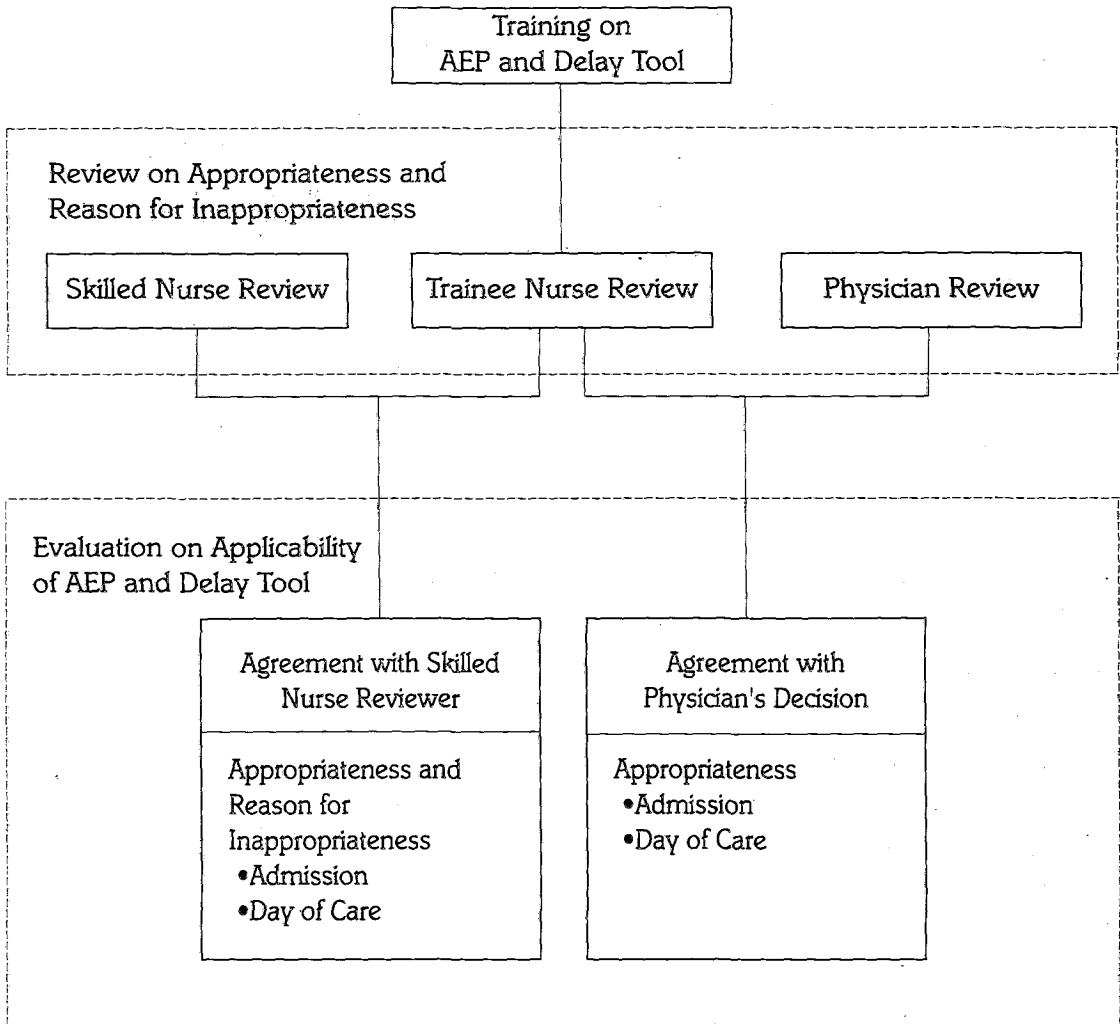


Fig. 1. Study framework

약 830병상 규모의 대학병원이며, 평균 재원일수는 약 10일, 병상 가동률은 약 87% 정도이다. 이 조사 대상 병원에서의 '병상 이용의 적절성 평가'의 조사 대상 중 일부를 다시 표본 추출하여, 피교육자와 숙련검토자 및 임상 의사간의 판단의 일치도를 평가하고자 하였다.

피교육자와 숙련검토자 및 임상 의사간의 판단의 일치도 평가의 조사 대상 환자 수와 재원일수는 Table 1과 같다. 내과, 소아과, 일반외과의 환자 52명과 104일 재원일에 대하여 각 검토자가 입원 및 재원일 적절성과 부적절 입원 및 재원의 이유를 조사

하였다. 단 임상 의사의 검토 시 의무기록을 구할 수 없었던 2례는 임상 의사의 조사 대상에서 제외되었다.

내과의 경우 '병상 이용의 적절성 평가'의 조사 대상 환자 중 50%를 계통적 추출(systematic sampling)하였고, 소아과와 일반외과에서는 조사 대상 환자 전수를 대상으로 하였다. 진료과별로 일정수 이상의 환자 및 재원일 조사 대상으로 선정함으로써, 검토자간의 일치도를 평가하는 데 진료과의 의학적 특성이 반영될 수 있도록 하였다.

Table 1. Number of patients and patient days by department

(Unit: patients, patient days)

Department	Patients		Patient days	
	N	%	N	%
Internal medicine	19(18)	36.5( 36.0)	38( 36)	36.5( 36.0)
Pediatrics	11(11)	21.2( 22.0)	22( 22)	21.2( 22.0)
General surgery	21(20)	40.4( 40.0)	42( 40)	40.4( 40.0)
Total	52(50)	100.0(100.0)	104(100)	100.0(100.0)

\* Note: in parenthesis, patients and patient days reviewed by physician.

### 3. 조사 도구

#### 1) 적절성 평가지침

적절성 평가지침은 입원 환자를 대상으로 입원 및 재원, 수술의 의학적 필요성을 판단하는 도구로 Gertman과 Restuccia<sup>18)</sup>에 의하여 개발되었다. 이는 각각의 항목에 대하여 적절성을 판단하는 기준들인 입원 기준과 재원 기준, 수술 기준으로 구성되어 있다.

적절성 평가지침에는 다음과 같은 특성이 있다.<sup>19)</sup>

1) 적절성 평가지침은 진단 명에 무관하게(diagnosis-independent) 적용할 수 있다. 2) 적절성 평가지침은 의학적 필요에 근거하여 적절성 여부를 판단하기 때문에 사회적 조건이나 병원 환경 등과 같은 요인은 적절성 판단에 영향을 미치지 않는다. 3) 적절성 평가지침은 입원 의료이용의 적절성을 평가하는 다른 의료이용검토(utilization review) 도구들에 비하여 높은 신뢰도와 타당도를 지니고 있다. 4) 적절성 평가지침은 의사가 하던 의료인 이라도 조사할 수 있도록 개발되었다.

#### 2) 이유목록(Delay Tool)

이유목록은 부적절 재원의 이유를 파악하기 위한 도구로 Selker 등<sup>4)</sup>에 의하여 개발되었으며, 우리 나

라의 상황을 반영한 이유목록 수정 판의 적용 가능성이 검토된 바 있다.<sup>6)</sup> 이유목록은 적절성 평가지침을 적용하여 부적절 입원 및 재원으로 판단된 경우, 미리 작성된 목록에서 해당하는 이유를 찾도록 되어 있다. 이유목록은 크게 부적절 입원의 이유와 부적절 재원의 이유로 구분되어 있다. 부적절 입원의 이유는 모두 4개 항목으로 이루어져 있으며, 부적절 재원의 이유는 이를 포괄적으로 정의하는 6개의 주범주와 각 주범주에 대하여 부적절 재원의 이유를 보다 구체적으로 기술하는 중범주와 소범주로 구성되어 있다.<sup>16)</sup>

### 4. 조사 방법

#### 1) 검토자

피교육자 간호사 2인, 숙련검토자 간호사 1인과 그리고 임상 의사 1인이 각각 독립적으로 적절성 및 부적절성의 이유를 판단하였다. 숙련검토자 간호사와 임상 의사는 각각 조사 대상 모두를 검토하였으며, 피교육자 간호사 중 1인은 내과 환자 및 재원일을 다른 1인은 소아과와 일반외과의 환자 및 재원일을 검토하였다.

피교육자인 간호사 2인은 간호사 경력 10년 이상이며, 적절성 평가지침과 이유목록에 대한 교육을

Table 2. Agreement on appropriateness between trainee and skilled nurse reviewer

(Unit: %)

Department	Admission review			Day review		
	Agreement		Kappa	Agreement		Kappa
	Overall	Specific		Overall	Specific	
Internal medicine	100.0	100.0	1.00*	94.7	92.9	0.87*
Pediatrics	100.0	-	-	100.0	-	-
General surgery	100.0	100.0	1.00*	100.0	100.0	1.00*
Total	100.0	100.0	1.00*	98.1	94.1	0.96*

\* p &lt; 0.01

받기 이전에는 이러한 의료이용검토 도구를 접한 적이 전혀 없었다. 숙련검토자 간호사는 적절성 평가 지침과 이유목록을 적용한 입원 및 재원일 적절성 평가를 직접 시행한 경험이 있고, 본 연구진이 이들 도구의 세부적인 지침을 마련하는 데 참여했었다. 임상적 판단은 5년 간의 임상경험을 가지고 있고, 적절성 평가지침의 판단 기준에 대해 사전 지식이 없는 가정의학 전문의에 의해서 이루어졌다.

## 2) 자료 수집 방법

본 연구에서는 환자 재원 시 의무기록을 검토하는 입원 중 검토(concurrent review)의 방법으로 입원 및 재원일 적절성과 부적절성의 이유를 평가하였다. 본 연구의 검토자들은 조사 대상 병원에서 '병상 이용의 적절성 평가조사'를 시작한 지 1주일이 되는 시점에 조사 당일과 조사 일 1일전, 모두 2일에 대하여 적절성과 부적절성의 이유를 평가하였다. 단 임상 의사 검토자는 부적절성의 이유에 대해서는 판단하지 않았다.

본 연구에서 사용한 입원 및 재원일 적절성 판단 기준은 근본적으로 적절성 평가지침의 원 개발자인 Gertman과 Restuccia의 판단기준<sup>20)</sup>과 동일하였다. 단, 본 연구진이 적절성 평가지침과 이유목록을 적용하였던 조사 경험을 활용하여, 적절성 평가지침과 이유목록의 각 기준에 대하여 세부적인 지침을 덧붙

였고 이를 적절성을 판단하는 데 활용하였다.

## 5. 일치도 평가 방법

검토자간의 일치도를 평가하는 지표로는 전반적 일치도(overall agreement)와 부적절성 판단 일치도(specific nonacute agreement), 카파 통계량(kappa statistics)을 사용하였고,<sup>3),17)</sup> 카파 통계량에 대하여 유의수준 5%에서 유의성 검정을 시행하였다.<sup>21)</sup>

피교육자의 적절성 평가지침에 의한 판단과 임상적 판단이 일치하지 않는 환례에 대해서는 검토자들과 연구진이 환례를 구체적으로 기술함으로써, 판단이 일치하지 않은 이유를 밝히고자 하였다.

## III. 연구 결과

### 1. 숙련검토자와의 일치도

#### 1) 입원 및 재원일 적절성

피교육자 간호사와 숙련검토자 간호사와의 입원 및 재원일 적절성에 대한 판단의 일치도는 Table 2와 같다. 입원 및 재원일 적절성에 대한 판단의 일치도는 매우 높은 수준이었다. 입원 적절성에 대한 판단의 전반적 일치도와 부적절성 판단 일치도는

Table 3. Agreement on reason for inappropriateness between trainee and skilled nurse reviewer

(Unit: %)

Department	Admission review			Day review		
	Number of observation	Overall agreement	Kappa	Number of observation	Overall agreement	Kappa
Internal medicine	16	93.8	0.00	26	88.5	0.81*
Pediatrics	0	-	-	0	-	-
General surgery	9	100.0	1.00*	6	100.0	1.00*
Total	25	96.0	0.88*	32	90.6	0.85*

\* p < 0.01

100%였고, 카파 통계량은 1.00이었다. 재원일 적절성에 대한 전반적 일치도 98.1%, 부적절성 판단 일치도 94.1%였고, 카파 통계량은 0.96이었다. 입원 적절성 및 재원 적절성에 대한 카파 통계량은 모두 통계적으로 유의하였다.

진료과별로는 내과에서의 재원일 적절성에 대한 판단의 일치도가 다른 진료과에 비하여 낮았다. 내과에서 재원일 적절성에 대한 전반적 일치도는 94.7%였고, 부적절성 판단 일치도 92.9%, 카파 통계량은 0.87이었다.

## 2) 부적절 입원 및 재원의 이유

피교육자 간호사와 숙련검토자 간호사의 부적절 입원 및 재원의 이유에 대한 판단의 일치도 Table 3과 같다. 부적절성의 이유에 대한 판단의 일치도 역시 높은 수준이었으나, 입원 및 재원일 적절성에 대한 판단의 일치도 수준보다는 낮았다. 부적절 입원의 이유에 대한 전반적 일치도는 96.0%, 카파 통계량은 0.88였으며, 부적절 재원 이유의 주범주에 대한 전반적 일치도 90.6%, 카파 통계량은 0.85였다. 부적절 재원 이유의 주범주가 일치하는 경우, 중범주도 모두 일치하였다. 부적절 입원 및 재원의 이유에 대한 판단의 카파 통계량은 모두 통계적으로 유의하였다.

진료과별로는 재원일 적절성의 경우와 마찬가지로 내과에서의 부적절성의 이유에 대한 판단의 일치도가 다른 진료과에 비하여 낮았다. 부적절 입원의 이유에 대한 판단의 전반적 일치도는 93.8%였고, 부적절 재원 이유에 대한 판단의 전반적 일치도는 88.5%였다.

## 2 임상적 판단과의 일치도 및 불일치의 이유

### 1) 일치도

피교육자 간호사가 적절성 평가지침을 적용하여 판단한 결과와 임상 의사 검토자가 묵시적 기준에 의해 입원 및 재원일 적절성을 판단한 결과간의 일치도는 Table 4와 같다. 적절성 평가지침의 판단과 임상적 판단간의 일치도는 높은 수준이었다. 그러나 이는 간호사 검토자들간의 일치도에 비해서는 낮은 수준이었다. 입원 적절성 평가에 대한 전반적 일치도가 86.0%, 부적절성 판단 일치도가 74.1%였고, 재원일 적절성 평가에 대한 전반적 일치도는 87.0%, 부적절성 판단 일치도는 69.0%였다. 이들에 대한 카파 통계량은 모두 0.71로 통계적으로 유의하였다.

진료과별로는 전반적 일치도와 카파 통계량은 피교육자 간호사와 숙련검토자 간호사간의 판단의 일치도의 경우와 마찬가지로 내과에서 일치도가 가장

Table 4. Agreement on appropriateness between trainee nurse reviewer and physician reviewer

(Unit: %)

Department	Admission review			Day review		
	Agreement		Kappa	Agreement		Kappa
	Overall	Specific		Overall	Specific	
Internal medicine	77.8	75.0	0.37	80.6	77.4	0.49*
Pediatrics	100.0	-	-	100.0	-	-
General surgery	85.0	72.7	0.70*	85.0	45.5	0.54*
Total	86.0	74.1	0.71*	87.0	69.0	0.71*

\* p &lt; 0.01

Table 5. Content analysis of disagreement between trainee nurse reviewer and physician reviewer

<b>Admission Review</b>
Physician reviewer: Necessary, Trainee nurse reviewer: Unnecessary - 3 admissions
1. Admission for pleural biopsy (Internal medicine, 2 cases)
2. Arteriovenous fistula infection (Internal medicine, 1 case)
Physician reviewer: Unnecessary, Trainee nurse reviewer : Necessary - 3 admissions
1. Admission for incision and drainage of peristomal abscess (General surgery, 1 case)
- operation was done in operating room
2. Admission for excision of breast mass (General surgery, 1 case)
- operation was done in operating room under general anesthesia
3. Gastrointestinal bleeding (General surgery, 1 case)
- Continuous infusion of intravenous fluid containing antacids
<b>Day review</b>
Physician reviewer: Unnecessary, Trainee nurse reviewer: Necessary, - 12 patient days
1. Carbon monoxide intoxication (Internal medicine, 2 patient days)
- Continuous infusion of intravenous fluid containing curmmulin and antibiotics
- NPO state
2. Fibrothorax pneumonia (Internal medicine, 2 patient days)
- Continuous infusion of intravenous fluid containing aminophylline
3. Hepatoma with ascites (Internal medicine, 2 patient days)
- Intravenous injection of Diuretics
4. Liver cirrhosis (Internal medicine, 1 patient days)
- Aldactone dose adjustment under medical supervision
5. Incision and drainage of peristomal abscess (General surgery, 2 days)
- Intravenous injection of antibiotics
6. Acute appendicitis (General surgery, 2 patient days)
- Intravenous injection of antibiotics
7. Duodenal ulcer with melena (General surgery, 2 patient days)
- Continuous infusion of intravenous fluid containing antacids, and packcell transfusion
- NPO state



낮았다. 그러나 부적절성 평가 일치도는 일반외과에서 가장 낮았다.

## 2) 불일치의 이유

피교육자 간호사의 적절성 평가지침에 의한 판단과 임상적 판단과의 일치하지 않는 입원 및 재원일을 기술한 내용은 Table 5와 같다.

입원 적절성의 경우 적절성 평가지침에 의한 판단과 임상적 판단과의 일치하지 않는 입원례는 내과 3건, 일반외과 3건으로 모두 6건이었다. 이 중 적절성 평가지침의 입원 기준을 만족하지 못하였으나 임상적으로 입원이 필요하다고 판단된 환례가 3건이었고, 반대로 적절성 평가지침의 기준은 만족하나 임상적으로 입원이 불필요하다고 판단된 환례가 3건이었다. 전자의 사례들 중 흉막 생검의 경우 적절성 평가지침 입원 기준 가운데 일부 기준을 적용할 수 있는 것으로 나타났으나 본 연구진이 마련한 세부 기준과 일치하지 않았다. 후자의 경우는 적절성 평가지침의 기준을 만족하나 번복기준(override option)을 사용할 수 있는 사례들로 판단되었다.

재원 적절성 판단에 있어서 적절성 평가지침에 의한 판단과 임상적 판단과의 일치하지 않는 재원일은 내과 7일, 일반외과 6일로 모두 13 재원일이었다. 이들은 모두 적절성 평가지침의 기준은 만족하나 임상적으로 재원이 불필요하다고 판단된 경우였다. 이를 재원일은 모두 의료서비스를 과잉이용(overutilization)한 경우로 판단되었으며, 따라서 번복 기준을 사용할 수 있는 사례로 판단되었다.

## IV. 고 찰

본 연구는 조사 대상 병원에서 '병상 이용의 적절성 평가'가 시작된지 일정 기간 후에 단면적 조사를 통하여 재원일 적절성 및 부적절 재원의 이유에 대한 검토자간의 일치도를 평가하였기 때문에, 조사 대상 재원일이 특정 재원 구간(part of stay)에 편중되었을 가능성이 있다. 재원 구간이 검토자간의 적

절성 판단의 일치도에 미치는 영향에 대해서는 아직 보고된 바 없지만, 연구 결과를 일반화하는 데 제한 요인으로 작용할 것으로 생각된다.

입원 중 검토 방법은 입원 및 재원일 적절성과 부적절성의 이유를 판단하는 데 있어서 의무기록 이외의 자료원을 활용할 수 있기 때문에 후향적 검토에 비하여 환자의 상황을 보다 정확하게 반영할 수 있다는 장점이 있다. 우리 나라의 경우 입원 의무기록의 충실도에 대한 평가는 없지만, 미국에 비하여 상대적으로 부실할 가능성이 높기 때문에 적절성과 부적절성의 이유에 대한 판단에 있어서 입원 중 검토의 방법이 검토자간 판단의 일치도를 높이는 데 기여했을 것으로 생각된다.

본 연구에서는 피교육자 간호사의 적절성 및 부적절의 이유에 대한 판단과 숙련검토자 그리고 임상 의사의 판단간의 일치도 수준을 평가함으로써 개별 병원에서의 적절성 평가지침과 이유목록의 적용 가능성을 평가하고자 하였다. 피교육자 간호사와 숙련 검토자 간호사간의 판단의 일치도 평가는 검토자간의 일치도(interrater reliability)에 가까우며, 피교육자 간호사의 이들 도구를 적용한 판단과 임상적 판단과의 일치도는 조사 도구의 수렴 타당도(convergent validity)에 가깝다고 할 수 있다.<sup>3)</sup> 그러나 입원 및 재원의 적절성에 대한 임상적 판단이 의사 1인에 의해서 이루어졌기 때문에 판단자의 주관적 요소가 배제되었다고 할 수 없다. 따라서 향후 적절성 평가지침의 적용 가능성을 보다 엄밀하게 평가하기 위해서는, 다수의 임상 의사가 입원 및 재원일의 적절성을 평가하도록 함으로써 검토자의 주관적 요소를 배제해야 할 것이다.

본 연구에서 입원 및 재원일 적절성에 대한 피교육자 간호사 검토자와 숙련검토자 간호사간의 판단의 일치도는 기존 연구 결과에 비하여 매우 높은 수준이었다. 기존 연구에서 입원 적절성에 대한 전반적 일치도 수준은 약 89~94% 정도, 부적절성 판단 일치도는 31~86%였고, 재원일 적절성에 대한 전반적 일치도 수준이 약 98~86%, 부적절성 판단 일치도는 약 53~89%였다.<sup>2,4,6,9,13,16-18,22,23</sup> 또한 부적절 입

원 및 재원의 이유에 대한 피교육자 간호사 검토자와 숙련검토자 간호사간의 판단의 일치도도 역시 매우 높은 수준이었다. 국내외 기존 연구에서 부적절 재원의 이유에 대한 검토자간의 전반적 일치도는 각각 63%와 94%였다.<sup>5),16)</sup>

이러한 결과는 개별 병원에서 적절성 평가지침이나 이유목록의 적용 가능성이 매우 높음을 시사하고 있다. 기존 연구에서는 숙련검토자간의 일치도를 비교하였던 것에 반하여, 본 연구에서는 피교육자에게 단기간의 교육만을 실시한 후에 숙련검토자와의 판단 결과를 비교하였음을 고려할 때 이러한 일치도 수준은 매우 높다고 할 수 있다.

이러한 높은 일치도 수준은 도구 자체의 간결성, 교육의 효과 등에 기인한 것으로 추정되며, 이와 함께 본 연구진이 적절성 평가지침을 적용한 연구 및 조사를 수행하는 과정에서 얻은 경험도 영향을 미쳤을 것으로 생각한다. 본 연구진은 적절성 평가지침과 이유목록을 적용한 여러 차례의 연구 및 조사를 통하여 원래의 적절성 평가지침과 이유목록의 기준에 대한 세부 지침을 개발해 왔다.

간호사 검토자들은 적절성 평가지침을 적용하여 입원 및 재원일의 적절성을 판단하는 데 있어서 번복기준을 전혀 사용하지 않았다. 외국의 연구 결과에 의하면 번복기준의 사용 여부가 검토자간의 판단의 일치도에 미치는 영향은 일관된 결과를 보이지 않고 있다.<sup>9),13),17),19)</sup> 그러나 적절성 평가지침의 개발자들에 의한 연구 결과<sup>19)</sup>를 제외하면, 한 연구에서는 전반적 일치도가 5~6%, 부적절성 판단 일치도가 11~12% 낮아졌다고 보고하고 있고,<sup>13)</sup> 다른 한 연구에서는 번복기준을 많이 사용한 검토자에서 일치도 수준이 낮아졌다고 보고하고 있다.<sup>9)</sup> 따라서 본 연구의 간호사 검토자들이 번복기준을 사용하지 않았던 것이 일치도 수준을 높이는 효과를 나타냈을 것으로 판단된다.

입원 및 재원일 적절성에 대하여 피교육자가 적절성 평가지침을 적용하여 판단한 결과와 임상적 판단간의 일치도 수준은 간호사 검토자들간의 일치도 수준에 비하여는 낮았으나, 기존 연구<sup>4),13)</sup>의 일치도

수준에 비해서는 높았다. 기존 연구에서 적절성 평가지침과 임상적 판단간의 부적절성 판단 일치도 수준은 입원 적절성의 경우 39~71%였고, 재원일 적절성의 경우 59~72%였다. 이러한 결과는 간호사 검토자들간의 입원 및 재원일 적절성에 대한 판단의 일치도 평가에서와 마찬가지로, 개별 병원에서 적절성 평가지침이나 이유목록의 적용 가능성이 매우 높음을 시사하고 있다.

그러나 앞서 언급한 바와 같이 본 연구에서는 1인의 임상 의사가 입원 및 재원일 적절성에 대한 판단하였으나, 기존 연구<sup>4,13)</sup>에서는 여러 명의 임상 의사가 적절성을 판단하도록 함으로써 검토자의 주관적 요소를 배제하였기 때문에 일치도 수준이 본 연구 결과에 비하여 다소 낮아졌을 가능성이 있다.

적절성 평가지침의 판단과 임상적 판단이 일치하지 않았던 환례의 대부분은 의료서비스를 과잉이용했던 경우였다. 이는 적절성 평가지침이 환자의 진료에 대한 의학적 결정이 옳다는 전제에 근거하고 있기 때문에, 잘못된 의학적 판단으로 인하여 의료서비스를 과잉이용하는 경우를 찾아내기가 어렵기 때문이다. 이러한 의료서비스의 과잉이용으로 인하여 생겨나는 임상적 판단과의 불일치 문제는, 적절성 평가지침의 적용 시 번복기준을 적절하게 사용함으로써 어느 정도 해결될 수 있는 것으로 판단된다. 그러나 번복기준을 적용하면 검토자간의 일치도가 낮아진다는 기존 연구의 결과를 고려할 때, 번복기준의 사용을 권장하기 이전에 이에 대한 적절한 판단 지침이 개발되어야 할 것으로 생각된다. 적절성 평가지침의 판단과 임상적 판단이 일치하지 않았던 환례 중의 일부는 본 연구진에 의하여 개발된 적절성 평가지침의 세부 지침과 임상 의사의 검토자의 판단이 일치하지 않았던 경우였다. 이는 임상 전문가들의 합의 과정을 통하여 적절성 평가지침의 세부 지침이 보다 정교하게 수정되어야 필요가 있음을 시사하고 있다.

진료과들 가운데 내과 환자에서 적절성 및 부적절성의 이유에 대한 판단의 일치도 수준이 가장 낮았다. 이는 간호사 검토자들간의 일치도와 임상적

판단과 적절성 평가지침에 의한 적절성 판단간의 일치도 모두에서 같은 결과를 나타냈다. 이러한 결과는 내과의 경우 임상적 상황이 다른 진료과에 비하여 복잡하여 간호사 검토자가 적절성을 판단한 데 어려움이 있었기 때문인 것으로 추정된다. 그러나 외국에서의 기존 연구에서는 진료과가 일치도 수준에 영향을 미치지 않았다고 보고하고 있다.<sup>17)</sup>

숙련검토자 간호사에 의한 판단과 임상적 판단간의 일치도는 피교육자 간호사와 임상적 판단간의 일치도에 비하여 높은 수준이었다. 숙련검토자 간호사에 의한 판단과 임상적 판단간의 일치도는, 입원 적절성의 경우 전반적 일치도가 92.0%, 부적절성 판단 일치도가 85.2%였고, 재원일 적절성의 경우 전반적 일치도는 89.0%, 부적절성 판단 일치도는 72.5%였다. 이러한 일치도 수준의 차이는 피교육자 간호사가 적절성 평가지침이나 이유목록에 익숙해졌을 때, 임상적 판단과의 일치도가 일정 정도 향상될 수 있음을 시사하고 있다.

본 연구는 1개 병원을 연구 대상으로 하였기 때문에 연구 결과를 일반화하기는 어렵다. 이는 병원의 규모나 진료 수준에 따라 피교육자인 간호사의 경험이나 특성이 다를 수 있는데, 이러한 피교육자의 경험이나 특성이 적절성과 부적절성의 이유에 대한 판단에 영향을 미칠 것으로 예상되기 때문이다. 이는 적절성에 대한 검토 시 자료 원으로 활용할 수 있는 의료진의 경험이나 특성의 경우에도 마찬가지일 것이다.

또한 본 연구에서는 내과, 소아과, 일반외과만을 조사 대상으로 하였는데, 진료과별로 검토자간의 일치도 수준에 차이가 있음을 고려할 때, 본 연구의 일치도 수준을 다른 진료과에도 적용하기는 어려울 것으로 판단된다. 따라서 향후 적절성 평가지침과 이유목록의 적용 가능성을 보다 포괄적으로 평가하기 위해서는, 여러 수준의 의료기관과 여러 진료과를 조사 대상에 포함시켜야 할 것으로 생각된다.

## V. 요약 및 결론

피교육자 간호사와 숙련검토자 간호사간의 입원 및 재원 적절성에 대한 판단의 일치도는 매우 높은 수준이었다. 입원 적절성에 대한 판단의 전반적 일치도와 부적절성 판단 일치도는 100%였고, 재원일 적절성이 경우 전반적 일치도 98.1%, 부적절성 판단 일치도 94.1%였다. 피교육자 간호사와 숙련검토자 간호사간의 부적절 입원의 이유에 대한 전반적 일치도는 95.0%였으며, 부적절 재원 이유의 주범주 및 중범주에 대한 전반적 일치도는 90.6%였다.

피교육자 간호사의 적절성 평가지침에 의한 판단과 임상적 판단간의 일치도 역시 높은 수준이었다. 입원 적절성의 경우 전반적 일치도가 86.0%, 부적절성 판단 일치도가 74.1%였고, 재원일 적절성의 경우 전반적 일치도는 87.0%, 부적절성 판단 일치도는 69.0%였다.

적절성 평가지침과 이유목록은 병상 이용의 적절성을 평가하는 데 유용한 도구이다. 또한 이들 도구는 다른 도구들에 비하여 매우 간결하여 사용하기 쉽고, 간호사를 검토자로 활용할 수 있기 때문에 비용이 적게 든다는 장점이 있다. 본 연구의 결과는 개별 병원에서 적절성 평가지침과 이유목록을 적용하여 병상이용의 적절성 수준을 모니터 하는 데 있어서, 기존 간호사 인력을 검토자로 활용할 수 있음을 시사하고 있다.

향후 국가적 차원에서 또는 개별 병원 단위에서 이들 도구를 적용하여 지속적으로 병상 이용의 적절성 수준을 모니터 함으로써 제한된 병상을 효율적으로 사용하는 데 기여할 수 있을 것이다. 또한 의료보험 보수지불제도가 현행 행위별 수가제(fee for service)에서 DRG 등을 이용하는 포괄수가제도로 바뀐다면, 적절성 평가지침과 이유목록과 같은 도구는 그 중요성이 더욱 커질 것으로 예상된다.

## VI. 참고문헌

1. Wakefield DS, Pfaller M, Hammons GT, et al. Use of the Appropriateness Evaluation Protocol for estimating the incremental costs associated with nosocomial infections. *Med Care* 1987; 25(6): 481-88.
2. 이상일, 김용익, 신영수. 병상 이용의 적절성에 영향을 미치는 요인에 관한 연구 -한 대학병원 자료를 중심으로-. *보건행정학회지* 1993; 3(1): 1-24.
3. Payne SMC. Identifying and managing inappropriate hospital utilization: a policy synthesis. *HSR* 1987; 22(5): 709-69.
4. Strumwasser I, Paranjpe NV, Ronis DL, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medirew Instrument, and Intensity-Severity-Discharge criteria. *Med Care* 1990; 28(2): 95-111.
5. 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(2): 112-29.
6. Apolone G, Alfieri V, Braga A, Caimi V, Cestari C, Crespi V, et al. A survey of the the necessity of hospitalization day in an Italian teaching hospital. *Quality Assurance in Health Care* 1991; 3(1): 1-9.
7. Baneres J, Alonso J, Broquetas J, Anto JM. Inadequate admission and inactive days of hospital stay in patients with chronic pulmonary disease and lung neoplasm. *Med Clin Barc* 1993; 100(11): 407-11(Abstract).
8. Davido A, Nicoulet I, Levy A, Lang T. Appropriateness of admission in and emergency department and cause of failure. *Quality Assurance in Health Care* 1991; 3(4): 227-34.
9. Eggimann BS. Hospital utilization review under field conditions: Potential and improvements. *Int J of Tech Ass* 1993; 9(4): 514-21.
10. Formby DJ, McMullin ND, Danagher K, Oldman DR. The Appropriateness Evaluation Protocol: Application in an Australian children's hospital. *Aust Clin Rev* 1991; 11(4): 123-31(Abstract).
11. Gloor JE, Kisson N, Joubert GI. Appropriateness of hospitalization in a Canadian pediatric hospital. *Pediatrics* 1993; 91(1): 70-4.
12. Mozes B, Schiff E, Modan B. Factors affecting inappropriate hospital stay. *Quality Assurance in Health Care* 1991; 3(3): 211-7.
13. Smith HE, Sheps S, Matheson DS. Assessing the utilization of in-patient facilities in a Canadian pediatric hospital. *Pediatrics* 1993; 92(4): 587-93.
14. Zwarenstein MF, Benatar SR, Shaw LR. The pattern of care in the medical wards of a teaching hospital. *S Afr Med J* 1990; 77(9): 460-3(Abstract).
15. 신의철, 맹광호. 대학병원 내외과 환자 재원일의 적정성 평가 및 관련 요인 분석. 가톨릭대학 의학부 논문집 1991; 44(1): 57-65.
16. 김윤, 이건설, 김창엽, 김용익, 신영수, 이상일. 부적절 재원의 이유. *예방의학회지* 1993; 26(4): 650-60.
17. Rishpon S, Lubacsh S, Epstein LM. Reliability of a method of determining the necessity for hospitalization days in Israel. *Med Care* 1986; 24(3): 279-82.
18. Gertman PM, Restuccia JD. The Appropriateness Evaluation Protocol: A technique for assessing unnecessary days of hospital care. *Med Care* 1981; 14(8): 855-71.

19. 서울의대 의료관리학교실, 한국의료관리연구원. 병상이용의 적절성 평가지침서. 서울, 1993.
20. Utilization Management Associates, Inc. Appropriateness Evaluation Protocol Reviewers' Manual. Utilization Management Associates, Inc., 1991.
21. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed., New York: John Wiley & Sons, 1981: 212-36.
22. Kemper KJ. The reliability and validity of pediatric appropriateness evaluation protocol. QRB 1989; 15(3): 77-80.
23. Kreger BE, Restuccia JD. Assessing the need to hospitalize children: pediatric appropriateness evaluation protocol. Pediatrics 1989; 84: 242-7.