

전자의무기록을 이용한 욕창발생 예측 베이지안 네트워크 모델 개발

조인숙¹ · 정은자²¹인하대학교 의과대학 간호학과 부교수, ²분당서울대학교병원 간호팀장

Predictive Bayesian Network Model Using Electronic Patient Records for Prevention of Hospital-Acquired Pressure Ulcers

Cho, In Sook¹ · Chung, Eunja²¹Associated Professor, Department of Nursing, Inha University, Incheon²Director, Department of Nursing, Seoul National University Bundang Hospital, Seongnam, Korea

Purpose: The study was designed to determine the discriminating ability of a Bayesian network (BN) for predicting risk for pressure ulcers. **Methods:** Analysis was done using a retrospective cohort, nursing records representing 21,114 hospital days, 3,348 patients at risk for ulcers, admitted to the intensive care unit of a tertiary teaching hospital between January 2004 and January 2007. A BN model and two logistic regression (LR) versions, model-I and -II, were compared, varying the nature, number and quality of input variables. Classification competence and case coverage of the models were tested and compared using a threefold cross validation method. **Results:** Average incidence of ulcers was 6.12%. Of the two LR models, model-I demonstrated better indexes of statistical model fits. The BN model had a sensitivity of 81.95%, specificity of 75.63%, positive and negative predictive values of 35.62% and 96.22% respectively. The area under the receiver operating characteristic (AUROC) was 85.01% implying moderate to good overall performance, which was similar to LR model-I. However, regarding case coverage, the BN model was 100% compared to 15.88% of LR. **Conclusion:** Discriminating ability of the BN model was found to be acceptable and case coverage proved to be excellent for clinical use.

Key words: Pressure ulcer, Bayesian prediction, Logistic models, Risk assessment, Data mining

서 론

1. 연구의 필요성

입원기간 중 발생하는 욕창은 미국 의료연구 및 품질기관 (AHRQ)에서 강조하는 환자안전 지표 중 하나이다. 욕창은 미국에서만 2006년 322,946명의 메디케어 환자에게 발생하여 연 110-130억 불을 소비한 고비용의 유해사례로, 그동안 관련 연구들이 많이 이루어져왔다. 그러나 의료기관에서 보고한 욕창 발생은 1993년 이후 80% 가까이 증가하는 등 만족할만한 수준에 이르지 못하고 있다

(Russo, Steiner, & Spector, 2008). 욕창과 관련된 연구는 크게 세 가지 방향으로 정리할 수 있다. 하나는 다양한 임상 환경에서 위험요인 또는 예측요인을 찾아내는 연구(Baumgarten et al., 2008; Fogerty et al., 2008)이며, 다른 하나는 Braden 도구나 Norton 도구와 같은 위험 사정도구 정확도를 측정, 비교하는 연구이다(Defloor & Grypdonck, 2005; Hatanaka et al., 2008). 이러한 연구들의 주요 관심은 주어진 환경에 적합한 도구를 선정하거나 수정하는 것, 그리고 그러한 도구의 변별력을 높일 수 있는 기준점(cut-off)을 찾는 것이었다. 나머지 하나는 중재 개발 및 효과에 대한 연구이다(De Laat et al., 2007;

주요어: 욕창, 베이지안 네트워크 모델, 로지스틱 모델, 위험사정, 데이터 마이닝

*본 논문은 2009년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2009-0053032).

*This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. 2009-0053032).

Address reprint requests to: Cho, In Sook

Department of Nursing, Inha University, Yonghyeon 4-dong, Nam-gu, Incheon 402-751, Korea

Tel: +82-32-860-8201 Fax: +82-32-874-5880 E-mail: insook.cho@inha.ac.kr

투고일: 2010년 7월 16일 심사회의일: 2010년 7월 21일 게재확정일: 2011년 6월 7일

Jones, Fennie, & Lenihan, 2007).

위험사정 도구를 실무에 적용하는 데 있어서는 몇 가지 문제점이 있다. 그중 하나는 측정의 정확성, 의료팀 내 측정자 간 신뢰도, 측정의 일관성 문제이다(Beekman et al., 2007; Brown, 2004). 이는 도구 정확성에 직접적으로 영향을 주는 요인으로 사용자 대상의 주기적인 반복 교육이 필수적인 부분이다. 하지만 대부분 실무 환경은 잦은 간호사 교체, 이직 등 급속히 증가하는 교육 요구를 따라가지 못하고 있는 실정이다. 일례로 Braden 도구의 실무 유용성을 조사한 연구(Cho & Noh, 2010)에 따르면, 예측도구로서의 전반적 변별력이 기대 수준에 못 미치는 것으로 나타났다. 위험사정 도구를 전향적으로 연구한 Schoonhoven 등(2002) 또한 위험사정 도구의 임상 변별력이 낮다는 결과를 제시하였다. 이는 그동안 알려진 Braden 타당도 연구 결과(Bergstrom, Braden, Kemp, Champagne, & Ruby, 1998)와 차이를 보이는 것으로 실무 현장의 다양성과 복잡성, 사용자 요인이 도구 사용 정확성에 크게 영향을 주고 있다는 것을 의미한다. 또 다른 문제점으로는 수작업의 반복 적용이 필수적이라는 것이다. 간호사는 전체 환자를 대상으로 매일 또는 일정 주기로 도구를 사용한다. 이러한 불특정 다수를 대상으로 하는 반복 작업은 도구의 실무 활용을 저해하는 요인이 되고 있다.

그래서 Hatanaka 등(2008)은 전자의무기록의 진단검사의학 검사 결과만을 이용한 욕창 예측모델이 Braden 도구보다 예측력이 높음을 보여주는 새로운 접근을 소개하였다. 전자의무기록은 검사결과 외에 간호관찰과 사정, 활동 등 간호실무 데이터도 포함하고 있다. 이러한 데이터는 획득과 동시에 추적이 가능하며 기록 목적 이외 간호사의 실시간 임상 의사결정지원뿐 아니라 임상 연구에 활용될 가능성을 갖고 있다(Pesut, 2006). 그래서 의료정보 분야에서는 오래 전부터 임상 의사결정지원 시스템 개발에 간호실무 데이터를 이용해왔다. 임상 의사결정지원과 관련해서는 전통적으로 통계기법을 널리 이용해왔으나 인공지능분야의 기술이 발달하면서 의료에도 다양하게 적용되고 있다. 그중 베이지안 네트워크(Bayesian network, BN)는 베이지안 확률이론을 기반으로 생리현상, 질병과정, 업무 프로세스와 같이 기존 지식만으로 인과관계가 완전히 설명되지 않는 불확실성이 내재된 의료 분야 추론에 장점을 갖고 있다(Lucas, van der Gaag, & Abu-Hanna, 2004). BN은 간호분야에도 소개되어 간호연구에 유용하게 활용될 가능성이 제시되어 왔으나, 아직 구체적인 실무 적용 사례는 찾아보기 어려운 실정이다(Cho, Yoon, Park, & Lee, 2008; Lee, Abbott, & Johantgen, 2005).

이에 본 연구에서는 욕창발생 예측에 BN 모델 적용 가능성을 탐색하고자 하였다. 즉 간호사의 환자사정, 중재와 진단검사 같은 임상 데이터로부터 확률이론을 이용하여 욕창발생 예측모델을 살펴 보았다. 이를 위해 욕창발생 빈도가 높은 중환자실 환자를 대상으

로 했으며, 모델에 대한 평가는 예측 타당성뿐만 아니라 기존 연구에서 많이 사용된 회귀분석 기법과 예측력을 비교함으로써 BN 모델의 상대적 유용성도 살펴보았다. 또한 BN모델의 사례(case) 커버리지를 통해 실무 적용 가능성도 평가하였다.

구체적 연구 목적은 다음과 같다.

첫째, 간호실무 데이터를 이용해 욕창발생 예측 BN모델과 로지스틱 회귀(logistic regression, LR)모델을 개발한다.

둘째, BN모델과 LR모델의 예측 정확성, 변별력, 그리고 실무 적용 가능성을 비교한다. 정확성과 변별력은 민감도, 특이도, 양성 예측도, 음성 예측도, 승산비, area under the receiver operating characteristic (AUROC) 지표를 이용하고, 각 모델의 실무 적용 가능성은 사례 커버리지(case coverage)를 이용한다.

연구 방법

1. 연구 설계

전자의무기록으로부터 후향적으로 익명화하여 수집한 환자의 인구학적 자료와 간호실무 자료를 이용한 후향적 코호트(retrospective cohort) 연구이다.

2. 연구 대상

일반적으로 후향적 데이터 마이닝에서는 대상 데이터베이스 전체를 활용하거나 대표성을 갖는 무작위 선택 표본을 사용한다. 전체 데이터가 이용 가능하고 현재 컴퓨팅 파워가 수용할 수 있는 수준이라면 전체 데이터를 활용하여 모델구축과 타당도 검사 세트를 독립적으로 구분한다(Fernandez, 2010). 이에 본 연구에서는 전자의무기록을 통해 중환자실 입원환자 데이터가 이용 가능한 경기도 소재 3차 의료기관을 대상으로 2004년 1월부터 2007년 1월까지 3년 동안 외과계 중환자실에 입원했던 환자 중 욕창발생 위험 성인 환자 3,348명을 선정하였다. 이 기간은 데이터 수집이 진행된 2008년을 기준으로 해당 의료기관 개원 후 전자의무기록 시스템 운영이 안정화된 2004년까지 후향적으로 선택하였다. 대상 환자의 선택 기준은 다음과 같다.

- 1) 만 18세 이상의 성인 환자
- 2) 응급실을 포함해 입원 후 총 재원일수가 3일 이상인 환자
- 3) 중환자실 체류기간이 2일 이상인 환자
- 4) 입원 시 그리고 중환자실 입실 시 작성된 '간호정보조사'와 '중환자실 간호기록'의 간호사정 섹션 욕창발생 유무 항목에 입력된 값이 '없음'이었던 환자

이러한 기준에 의해 선정된 환자는 중환자실 체류기간 중 욕창발생 유무에 따라 욕창 발생군(n=205)과 비발생 위험군(n=3,143)으로 구분하였다. 3,348명 대상자의 중증도는 중환자실 분류체계에 따라 각각 1군 0%, 2군 0.04%, 3군 1.23%, 4군 27.64%, 5군 68.22%, 6군 2.86%이었다. 중증도 분류에 사용된 중환자실 분류체계는 1992년 임상간호사회에서 개발된 도구로 APACHE III를 이용한 Lee 등 (2000)의 연구에서 타당도가 검증된 도구였다.

3. 연구 도구

1) 중환자실 전자의무기록시스템

본 연구 자료는 간호실무 데이터로서 전자의무기록 중 중환자실 간호기록과 '간호일지' 화면으로 입력된 데이터이다. 중환자실 간호 기록은 구조화된 데이터 입력을 지원하며, 입력된 데이터는 혈액학 상태, 의식 및 신경계, 호흡기, 섭취량 및 배설량, 기타 간호사정이라는 5개의 데이터 범주로 구분된다. 이 중 간호사정 범주는 일단 환자의 욕창발생 유무와 욕창부위, 욕창분류를 기록하는 구조화된 항목이 있다. 본 연구에서는 이 욕창발생 유무 값을 예측 모델의 표준값(standards)으로 사용하였다.

2) 베이지안 네트워크

데이터 마이닝은 대량의 임상 데이터베이스로부터 유용한 정보나 패턴을 찾아내는 지식 발견(knowledge discovery in databases)의 데이터 처리과정을 의미한다. 데이터 마이닝에는 인공지능, 의사결정나무, 규칙발견, BN 등의 다양한 모델을 사용하며 기존에 수집된 데이터로부터 후향적으로 특정 결과변수와 관련된성이나 변수 간 관련성을 통계적 기법을 통해 찾아나가는 연구방법이다(Theide & Sewell, 2010).

그중 BN은 의사결정 모델 중 하나로 이미 알려진 지식체를 이용해 '임상관찰-결과'의 인과관계 또는 연관관계로 지식을 구조화하고 알려진 사전확률을 결합시켜 사후확률을 추론하는 수학적 모델이다(Korb & Nicholson, 2004). BN은 변수와 변수들 간의 관계를 노드(node)와 아크(arch)라는 순환되지 않는 그래프 형태로 표현한다. 이 구조는 명시적인 상호연관성을 갖는 변수들의 결합 확률분포(joint probability distribution)를 표현하는 것으로 변수 간의 의존성에 따른 조건부 확률 테이블로 구성된다. BN은 그래프 형태로 시각화되어 표현되기 때문에 의료진과 해당 분야 전문가들이 이해하고 소통하는 데 큰 장점을 갖고 있다. BN을 모델링하는 데 이용할 수 있는 도구는 SPSS의 클레멘타인 10.1 (SPSS Korea), Hugin (Hugin Expert, Denmark), Dezide advisor (Dezide Aps, Denmark) 등이 있으며, 본 연구에서는 Netica v.3.24 (Norsys Software Corp. Vancouver,

CANADA)를 이용하였다.

4. 자료 수집 방법 및 자료 분석

1) 예측 모델을 위한 변수 선정

본 연구자는 사전 연구(Cho, Yoon, Park, & Lee, 2008)에서 이 과정을 자세히 기술한 바 있다. 따라서 여기에서는 변수 선정과정을 요약하여 소개하고자 한다. 모델 구축을 위한 변수 선택은 우선 문헌 고찰을 통해 욕창관련 개념을 찾고, 이를 실무 데이터와 의미적으로 매핑하고, 입원일 기준으로 해당 데이터를 추출, 요약하였다. 이 과정에서는 연구팀 이외 대상 중환자실에서 10년 이상 경력을 가진 간호사 5인이 참여하여 간호제공이나 기록 과정에서 해당 항목을 언제, 어떠한 의미로 사용하고 있는지, 그리고 의미적 매핑 과정이 적절한지 교차 확인하였다.

2) 전자의무기록 데이터 수집

연구 대상자 코호트는 해당 의료기관 임상데이터 저장소 (clinical data repository, CDR)에서 대상자 선정 기준 검색을 통해 추출하였다. Figure 1은 CDR로부터 환자 선정기준에 따라 3년간 중환자실 총 입원 19,756건 중 3,348건을 검색한 과정과 3,348 입원에 대한 분류 과정을 나타내고 있다. 본 연구는 해당 의료기관 연구윤리심의위원회 심사를 거쳐 승인 받았으며, 데이터 추출과정에서 환자를 식별할 수 있는 환자 인식 정보는 삭제하였다.

3) 예측 모델 구축 및 평가방법

예측 모델 구축 과정을 정리하면 Figure 2와 같다. 우선 예측변수를 찾기 위해 욕창 발생군과 비발생 위험군 간의 카이검정과 t-test를 수행하고, 유의수준 .05를 기준으로 유의한 차이가 있는 변수 36개를 이용해 BN모델과 LR모델을 구축하였다. 이 과정에서 LR모델의 경우 변수 누락률에 영향을 받는 점을 감안하여 두 가지 접근법을 이용하였다. 즉 50% 이상의 누락이 있는 변수를 제거하고 50% 이하 누락률을 보인 변수만 포함하는 LR모델-I과 10% 이하의 누락률을 보인 변수만을 포함하는 LR모델-II로 구분하였다. LR모델 구축 과정에서 사용된 변수 선택 방법으로는 우도비검정과 전진선택법을 이용하였다.

구축된 모델의 타당도 검증은 3배수 교차타당성(threefold cross-validation) 방법을 이용하였다. 즉 검증용 데이터를 3묶음으로 무작위 배정하여 돌아가면서 2/3 데이터는 모델 구축에 사용하고, 나머지 1/3은 구축된 모델을 평가하는 데 사용하였다. 데이터의 무작위 배정을 위해 무작위 숫자 발생 함수를 이용, 입원일별 8자리 이내 숫자를 생성한 뒤 정렬시켜 묶음을 만들고, 순차적으로 2:1 비율로

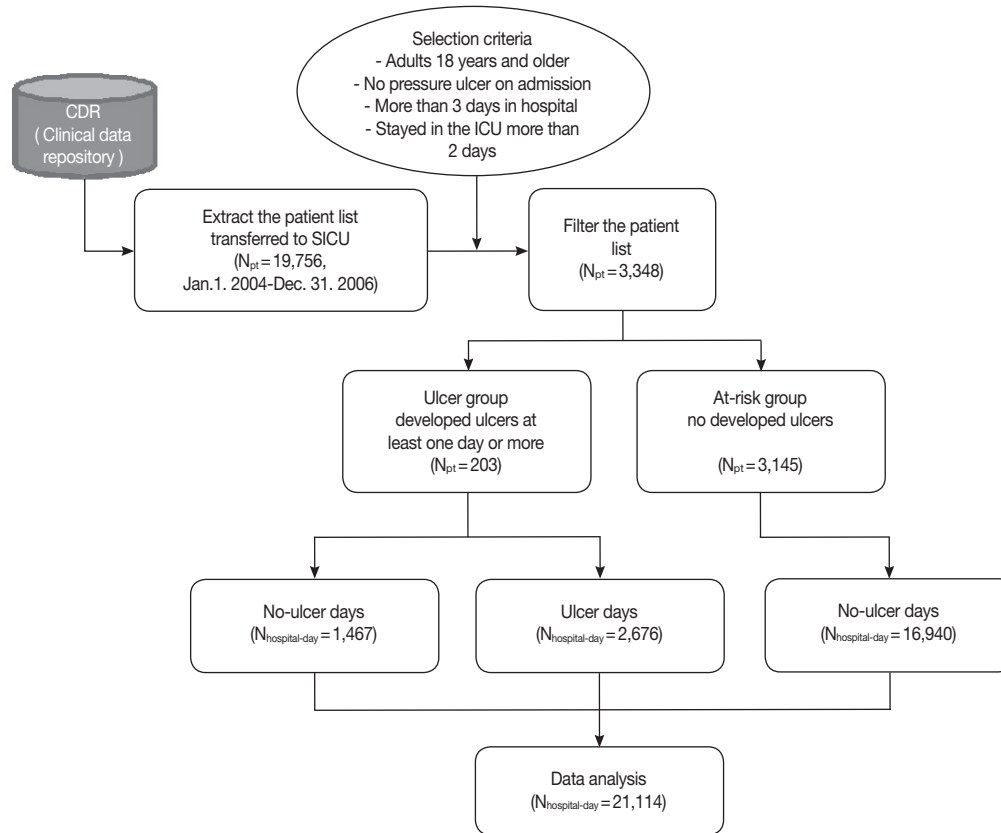


Figure 1. Procedure for extracting data from the clinical data repository (CDR) of the research hospital. N_{pt} and $N_{hospital-day}$ are the numbers of patients and hospital-day, respectively.

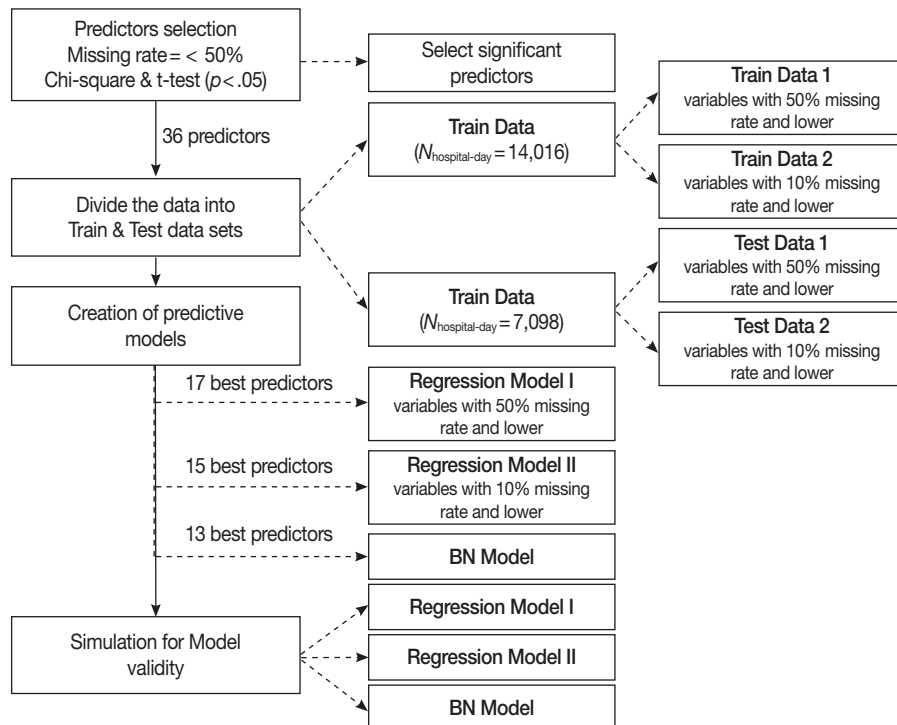


Figure 2. A summary of analytic procedures, $N_{hospital-day}$ is the number of hospital-day, BN stands for Bayesian Network.

모델 구축과 평가 데이터로 구분하여 3회 반복하면서 모든 묶음에 한 번씩은 평가 데이터로 지정되도록 하였다.

구축된 모델 평가는 연구설계 시 설정한 표준값과 예측값을 비교하여 민감도, 특이도, 양성예측도, 음성예측도, 승산비를 비교하였다. LR모델의 분류기준(cut-off)은 최적의 예측확률을 나타내는 값으로 LR모델-I의 경우 .12, LR모델-II의 경우 .14이었다. 그리고 모델의 예측력과 실무 적용 커버리지는 ROC 면적과 수식 (1)을 이용해 산출하였다. ROC 면적은 진단적 도구의 변별력을 나타내는 지표로 측정단위에 독립적이고, 도구 간 비교를 위한 범용적 척도가 되며, 다양한 기준값에 대한 가능한 성능 범위를 제시하기 때문에 예측이나 진단검사에 많이 활용되고 있다. 수식 (1)은 본 연구에서 정의한 사례 커버리지로서 전체 21,114 입원일 중 해당 모델 구축과 평가에 사용된 입원일 수에 대한 비율이다.

$$\text{Case coverage (\%)} = \frac{\text{hospital days used in model construction and evaluation}}{21,114 \text{ hospital days}} \times 100 \quad (1)$$

BN모델 변수 선정은 네트워크 구조 내에서 반응변수와 특정 변수 간의 상호정보량(mutual information)을 기반으로 하였다. 상호정보량은 정보이론에서 쓰이는 개념으로 사건 y에 대한 결과를 알고 있을 때 무작위 변수 x의 불확실성을 추정하는 개념이다(Korb & Nicholson, 2004). 이는 BN과 같은 추론 모델 구성 시 많은 데이터에서 무작위로 데이터 표본을 재구성할 때 생기는 불확실성을 줄이는 효과를 갖고 있어 유용하다. 본 연구에서는 Cover와 Thomas (1991)의 엔트로피라는 정보량 기대값을 이용해 수식 (2)와 같이 상호정보량을 산출하였다.

$$I(X;Y) = \sum_{xy} P_{XY}(xy) \log \frac{P_{xy}(x,y)}{P_X(x)P_Y(y)} = E_{P_{XY}} \log \frac{P_{XY}}{P_X P_Y} \quad (2)$$

모델선정은 36개 변수를 모두 이용해 구축한 BN구조를 바탕으로 포화 모델 민감도 분석을 수행하여 반응변수에 미치는 개별 변수들의 상호정보량과 구조 내 변동값을 이용하였다. 민감도 분석 결과 13번째 변수 이후 상호정보량이 <.0003 이하로 급속히 감소하는 것으로 나타나 이 지점을 기준으로 최종 13개 변수가 선택된 모델을 선정하였다.

연구 결과

1. 대상자의 일반적 특성 비교

수집된 입원일 데이터를 환자 기준으로 살펴보면 남성이 약 2/3,

Table 1. Group Comparison of Demographic Characteristics between Ulcer Group and Risk Group

Characteristics		N	Pressure-ulcer group (n=205)	At-risk group (n=3,143)	t or χ^2	p
Gender	M	2,016	140 (68.29%)	1,876 (59.69%)	5.95	.015
	F	1,332	65 (31.71%)	1,267 (40.31%)		
Mean age in years			63.1 (15.3)	59.0 (15.7)	3.59	<.001
Length of hospital stay			84.04 (79.83)	34.89 (50.30)	8.70	<.001
Length of stay in ICU			20.36 (25.93)	5.39 (9.60)	8.23	<.001
Medical diagnosis	Neurological	1,084	86 (41.95%)	998 (31.75%)	24.96	24.96
	Cardiovascular	750	44 (21.46%)	706 (22.46%)		
	Surgical	601	46 (22.44%)	555 (17.66%)		
	Respiratory	581	13 (6.34%)	568 (18.07%)		
	Others	332	16 (7.80%)	316 (10.05%)		

여성이 1/3을 차지하고 있었고, 욕창군 남성 비율이 위험군에 비해 약 10% 정도 많았다(Table 1). 평균 연령은 욕창군의 경우 약 63세인데 비해 위험군은 59세로 욕창군 연령이 유의하게 높았다. 병원 재원일 수는 욕창군이 위험군보다 약 2.5배 더 길었고 중환자실 체류일 수 또한 4배 정도 높은 것으로 나타났다. 진단명 분류를 살펴보면, 전체적으로 신경계, 심혈관계, 외과계, 호흡기계, 그 외 기타 순으로 많은 수를 차지하고 있었으며, 욕창군에서 신경계 환자가 약 10%, 외과계 환자가 약 5% 정도 더 많았다. 그에 비해 위험군은 호흡기계 환자 비율이 높아 군 간의 진단명 분류 분포가 유의하게 다른 것으로 나타났다. 욕창 발생률은 3,348명 중 205명으로 약 6.12%였다.

2. 욕창발생 예측 로지스틱 회귀모델

두 개의 LR모델 중 모델-I은 36개 변수 중 17개 변수가 선택되었다(Table 2). 모델-II의 경우는 변수 데이터 이용 가능성이 90% 이상인 변수 24개 중 성별, 입실경로, 진단명, 소변 회수, 대변 회수, 장루 여부, 유치도뇨 여부, 체위변경 회수 8개 변수가 모델-I과 동일하게 선택되었다. 그 외 연령, 입원일 수, 근육이완제 적용 여부, 비경구영양 여부, 진정 여부, 항응고제 투여 회수, 진정제 투여 회수의 7개 변수는 모델-I과 다르게 선택되었다. 두 모델의 회귀계수에 대한 검정은 Wald statistics에서 통계적으로 유의한 것으로 나타났으며, 모델에 포함된 변수 수에 따른 영향을 보정한 모델 적합도(Akaike Information Criterion)는 모델-I이 더 적합한 것으로 나타났다. 모델의 수정

Table 2. Multiple Logistic Regression Models Examining Risks of Pressure Ulcers

Predictors	Model-I		Model-II	
	Beta	OR (95% CI)	Beta	OR (95% CI)
Male	0.32	1.38* (1.06, 1.79)	-0.42	0.66† (0.58, 0.73)
Body mass index	-0.09	0.91† (0.87, 0.96)	-	-
Transfer from ER	-1.57	0.21† (0.13, 0.35)	-0.21	0.81* (0.74, 0.90)
Medical diagnosis				
Respiratory	Ref.	1.00	Ref.	1.00
Neurological	0.78	2.23* (1.27, 3.95)	-0.27	0.76† (0.63, 0.92)
Cardiovascular	0.62	1.86* (1.08, 3.22)	0.20	1.22† (1.02, 1.47)
Surgical	0.31	1.43 (0.79, 2.56)	-0.04	0.96† (0.79, 1.17)
Other	0.98	2.70* (1.47, 4.95)	-0.16	0.85† (0.68, 1.07)
Systolic blood pressure	0.02	1.02† (1.01, 1.03)	-	-
Diastolic blood pressure	-0.04	0.96† (0.94, 0.98)	-	-
Number of urinations	0.08	1.08† (1.05, 1.11)	0.11	1.11† (1.10, 1.12)
Number of defecations	0.08	1.08† (1.04, 1.13)	0.06	1.06† (1.05, 1.07)
Any stomy	-0.85	0.42† (0.29, 0.61)	-0.43	0.65† (0.59, 0.72)
Number of drain tubes	0.02	1.02† (0.06, 0.65)	-	-
Any catheterization	-1.59	0.20† (0.06, 0.65)	-1.30	0.27† (0.22, 0.35)
Number of position changes	0.18	1.20† (1.14, 1.26)	0.07	1.07† (1.05, 1.08)
Peripheral sensory: not intact	-0.32	0.67† (0.48, 0.92)	-	-
Skin condition: not intact	-1.02	2.76† (2.12, 3.59)	-	-
Albumin	-1.10	0.33† (0.21, 0.52)	-	-
Number of sedatives administered	-0.36	0.70† (0.53, 0.92)	-	-
Any strait	-0.32	0.72* (0.55, 0.93)	-	-
Age in years	-	-	0.01	1.02† (1.01, 1.02)
Length of stay	-	-	0.01	1.02† (1.02, 1.03)
Any inotropic support	-	-	-0.46	0.63* (0.48, 0.83)
Total parenteral nutrition	-	-	0.65	1.91† (1.72, 2.11)
Any sedation	-	-	-0.46	0.63† (0.53, 0.74)
Number of anticoagulants administered	-	-	-0.90	0.69* (0.60, 0.79)
Number of analgesics administered	-	-	-0.38	0.41† (0.25, 0.68)
Model-specific				
Wald χ^2 statistic		$\chi^2 [20] = 510.47$		$\chi^2 [18] = 2,401.96$
AIC		1,775.93		14,321.90
R^2_{adj}		0.40		0.24
Hosmer and Lemeshow goodness-of-fit test		$\chi^2 [8] = 8.91$		$\chi^2 [8] = 149.70$

*Confidence intervals (95% CI) not containing the null value (1.00) are statistically significant at $p < .05$; † $p < .001$ for χ^2 tests; The Akaike information criterion (AIC), a measure of statistical model fit, was used to compare the amount of information explained across the logistic regression models. A lower AIC value indicates a model is a better fit for the observed data.

ER=emergency room; OR=odds ratio; CI=confidence interval.

결정계수(R^2_{adj})를 보면 모델-I의 경우 .40, 모델-II는 .24이었다. 관찰된 확률과 모형에 의해 예측된 확률을 비교하는 Hosmer and Lemeshow 적합도 검증에서는 모델-I의 적합성은 지지되었으나($\chi^2 [8] = 8.91, p > .05$), 모델-II는 모형이 적합하지 않은 것으로 나타났다($\chi^2 [8] = 149.70, p < .001$).

3. 욕창발생 예측 베이지안 네트워크 모델

개발된 BN모델은 14개 변수 노드와 13개 링크, 1,957개 조건부 확

률 테이블로 구성되었다. 구성 변수에서 회귀모델과 다른 점은 중환자실 체류일 수, 식이, 의식수준, 체온, 자가 운동능력, 헤모글로빈 6개 노드가 추가된 것이었다. BN모델에 대한 민감도, 특이도, 양성 예측도, 음성 예측도, 승산비를 산출한 결과 Table 3에서 보듯이 10% 기준점에서 민감도 88.57%, 특이도 65.16%, 양성예측도는 29.49%, 음성예측도 97.2%를 보였으며, 20% 기준점에서는 각각 81.95%, 75.63%, 35.62%, 96.22%였다. 진단적 검사의 기준점 결정에 사용되는 (0, 1)지점 근접성과 Youden index (J)를 산출해보면 20%에서 (0, 1)지점 근접성이 가장 좋았고, J 지표도 가장 높게 나타났다. 예측

Table 3. Results of the Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value

Cut-off	Sensitivity	Specificity	PPV	NPV	Odds	95% CI
0	100	0	14.13	100	-	-
2	96.39	44.66	22.27	98.69	21.55	15.11-30.73
5	93.02	56.22	25.9	98	17.11	15.70-18.66
10	88.57	65.16	29.49	97.2	14.49	13.46-15.61
20	81.95	75.63	35.62	96.22	14.09	13.16-15.09
30	74.73	81.04	39.33	95.12	12.64	11.82-13.52
40	67.63	84.9	42.42	94.1	11.75	10.97-12.59
50	60.29	88	45.26	93.09	11.13	10.36-11.97
60	54.87	90.46	48.61	92.42	11.53	10.67-12.46
70	47.77	92.52	51.23	91.5	11.31	10.40-12.31
80	35.62	94.93	53.62	89.96	10.36	9.39-11.43
85	29.36	96.22	56.09	89.23	10.58	9.46-11.83
90	21.18	97.35	56.77	88.25	9.87	8.66-11.25
95	11.91	98.89	63.87	87.22	12.05	9.90-14.66
98	4.33	99.62	65.45	86.36	11.78	8.51-16.55
99.5	0.48	99.98	80	85.93	24.11	5.86-99.23
100	0	100	100	85.87	24.11	-

PPV=positive predictive value; NPV=negative predictive value.

모델의 변별력을 평가하는 AUROC분석 결과는 85.01%였다.

4. 예측 모델 비교 및 실무적용 가능성 평가

개발된 예측 모델 성능을 비교한 결과, Table 4에서 보는 바와 같이 민감도는 LR모델-I에 높았으며, 특이도와 양성예측도는 BN모델이 높게 나타났다. 음성예측도는 97% 전후로 모델 간 차이가 크지 않았으며, AUROC분석에서도 LR모델-I과 BN모델이 85% 전후 값을 보였다. 전체 데이터 세트 중 각 모델 구축과 평가에 사용된 사례 커버리지를 비교해보면, Table 5와 같이 LR모델-I의 경우 15.88% (3,352 일) 데이터가 사용되었으며 LR모델-II는 98.33% (20,762일), BN모델은 전체 입원일 데이터가 모두 사용되었다.

논 의

본 연구는 실무에 직접 적용할 수 있는 육창발생 예측 모델을 개발하는 연구이다. 일반적으로 예측 모델을 개발하는 목적은 해당 관심 현상이 발생하는 데 영향을 주는 요인을 밝히거나, 현상 결과 분류 목적을 위해 결과값을 미리 얻고자 할 때, 또는 새로운 사례에 대한 예측 결과 확률을 얻고자 할 때이다. 본 연구에서는 전자의무 기록으로 축적되는 간호실무 데이터를 이용하여 육창발생 가능성이라는 확률을 실시간 산출하여 간호사의 간호중재 계획에 도움이 되는 정보를 제공하고자 하는 것이었다. 그러한 접근법으로 본 연구는 베이지안 확률이론에 기반한 BN모델 적용 가능성을 모색하

Table 4. Test Characteristics of the Predictive Models

	Regression Model-I	Regression Model-II	Bayesian Network Model
Sensitivity	83.63%	76.13%	81.95%
Specificity	74.32%	74.51%	75.63%
Positive predictive value	30.73%	32.92%	35.62%
Negative predictive value	97.16%	95.04%	96.22%
Area under the ROC (95% CI)	84.53% (81.23-87.81)	80.72% (79.21-82.27)	85.01%

Table 5. Case Coverage by the Predictive Models

Data set	Number of record (%)			Total (%)
	Regression Model-I	Regression Model-II	Bayesian Network Model	
Training set	2,414	13,947	14,016	15,231(100.0)
Testing set	1,040	6,815	7,098	5,883 (100.0)
Total	3,352 (15.88)	20,762 (98.33)	21,114 (100.0)	21,114 (100.0)

기 위해 시도되었으며, 모델의 예측력 비교를 위해 LR모델을 이용하였다.

우선 육창 발생률을 살펴보면 본 연구에서는 발생률이 6.12%였다. 기존 국내 3차 의료기관에서 수행된 연구에 따르면 발생률은 26.1-31.3%로 높은 것으로 보고되어왔다(Lee, Jeong, & Jeon, 2003; Park, Park, & Yu, 2005). 그러나 최근 발표(Cho & Noh, 2010)에 따르면 중환자실 육창 발생률은 위험 환자수를 기준으로 5.9% 내외라고 하였다. 이러한 수치를 미국 보고서(Agency for Healthcare Research and Quality, 2009)에 나타난 급성의료기관 평균 육창 발생률 7%와 비교하면 유사한 수준인 것으로 해석할 수 있겠다. 특히 본 연구가 중증도가 높은 3차 의료기관 중환자실 환자를 대상으로 한 점을 고려하면 전체 의료기관 평균 발생률은 더 낮아질 것으로 보인다. 그러나 기존 국내 연구가 대부분 3차 의료기관에서 수행되었음에도 발생률 차이가 큰 것을 보면 의료기관 간의 변이(variation)가 있거나, 발생률 산출 방법에 차이가 있거나, 시간적 경과에 따른 간호서비스 향상에 의한 결과일 수 있다. 따라서 이러한 변이 또는 변화 유무와 변이에 미치는 영향에 대한 연구가 더 필요한 것으로 여겨진다.

연구에 포함된 3,348명 환자의 군별 일반적 특성을 살펴보면 남성에서 발생률이 높고, 연령이 많고, 재원기간이 길며, 신경계, 외과계 환자에게서 발생률이 더 높은 것으로 나타났다. 이러한 결과는 본 연구에서 선택한 변수가 문헌고찰을 통해 이미 육창발생과 관련이 있는 것으로 알려진 변수들이었기 때문에 예상할 수 있던 결과라고 할 수 있다. 한편 대상자의 일반적 특성 이외 39개 임상관찰과 생리적 변수를 이용한 모델링에서는 변수 값의 결측률(missing rate)이 회귀모델 변수선택에서 중요한 문제가 되었다. 즉 전체 변수 중 의식수준, 혈색소, 부종 정도의 3개 변수가 50% 이상, 그 외 14개 변수

가 10-49% 결측률을 보였다. 그에 따라 LR모델-I, II로 차별화하여 접근하였는데, 결측 변수가 많이 포함된 LR모델-I과 결측률이 낮은 변수만을 포함한 LR모델-II 간에 약 절반가량의 변수가 다르게 선택되었다. 공통적으로 선택된 변수로는 성별, 입실경로, 진단명, 소변 횟수, 대변 횟수, 장루여부, 유치도뇨 여부, 체위변경 횟수였다. 이중 성별, 입실경로, 진단명은 중환자실 입실시 일회 수집되는 데이터이며, 소변과 대변 횟수, 유치도뇨 여부, 체위변경은 매일 반복적으로 측정 또는 기록되는 변수라는 공통점을 갖고 있었다. 또한 모델의 적합도, 수정결정계수, 잔차분석 지표 모두 모델-I이 모델-II보다 적합한 것으로 나타났다.

BN모델의 경우는 결측률에 상관없이 모든 변수가 모델링에 사용되었으며 종속변수에 미치는 기여도에 따라 선택이 결정되었다. 선택된 변수는 약 절반에 해당하는 7개 변수가 회귀모델과 동일했으며 회귀모델에서 선택되지 않았던 중환자실 체류일 수, 식이, 의식수준, 체온, 자가 운동능력, 혈색소가 추가되었다. 추가된 변수 중 중환자실 체류일 수를 제외한 5개 변수는 결측률이 상대적으로 높은 변수들이었다. 그럼에도 불구하고 BN모델의 전반적 예측력을 의미하는 AUROC는 85% 이상으로 진단적 예측 모델로서는 중상위 수준의 정확도를 보여주었다(Fischer, Bachmann, & Jaeschke, 2003). 또한 욕창발생 위험군과 비위험군을 판단하는 기준 확률값으로는 민감도와 특이도 균형 면에서 10-20%가 가장 적절한 것으로 여겨진다. 이 범위에서 민감도는 80-90%, 특이도는 65-76% 수준으로 Braden 도구의 타당도 연구 중 가장 높은 정확도를 보고한 Bergstrom 등(1998)의 민감도 88%, 특이도 68%와 비교했을 때 유사한 수준이라고 볼 수 있다. 또한 양성예측도 30-36%는 의학의 유방암 검사 실무 표준 양성예측도인 25-40%와 비교할 때 유사한 수준으로 실무에서 충분히 수용 가능한 수준이라고 볼 수 있다(Burnside, Rubin, & Shachter, 2004).

개발된 모델의 예측력 지표를 비교해보면, 예측 정확도를 의미하는 지표들 간에 큰 차이가 없었다. 즉 LR모델-I과 BN모델은 민감도, 특이도, 양성예측도, 음성예측도, AUROC 모두 유사한 수준이었다. 그러나 각 모델 구축과 평가에 사용된 입원일 데이터 수를 비교해보면 BN모델의 경우 모든 입원일 데이터가 사용된 데 비해 LR모델-I은 약 16% 데이터만 이용되었다. 이것은 모델을 전향적으로 실무에 적용했을 때 BN모델은 모든 입원일에 대해 예측 확률값을 산출하지만 LR모델-I의 경우는 6일 중 하루 꼴로 예측값을 얻을 수 있다는 것을 의미한다. 즉 BN모델이 LR모델보다 실무 적용 가능성이 더 높다고 할 수 있다.

이처럼 환자기록을 통해 수집되는 임상 데이터는 변수에 따라 결측률이 다양하며 불완전한 특성을 갖고 있다. 이는 환자기록이 연구 목적의 설문조사와 달리 임상 상황이나 의료진의 의도에 따라

수집되기 때문에 동일한 조건의 환자로부터 획득되는 데이터는 종류와 형태, 시간이 다를 수 있다(Harrison, 2008). 따라서 임상 데이터를 이용하는 데 있어서 이러한 특성에 대한 고려가 필수적이며, LR모델은 변수선정이나 데이터 커버리지 측면에서 제약이 많다고 할 수 있다. 이에 비해 불확실성과 사전확률을 이용하는 BN모델은 LR모델이 갖고 있는 제약을 보완할 수 있는 방안으로 여겨진다. 그래서 의학분야에서는 질병의 진단, 예후 예측과 치료방법 선택에 임상 의사결정 모델로 BN모델이 많이 활용되고 있으며, 의료기관 정보화와 함께 환자문제 예측과 간호계획 등에도 활용이 기대되고 있다(Lucas, van der Gaag, & Abu-Hanna, 2004).

최근 국내의 의료계 전자의무기록 도입이 빠르게 확산되고 있다. 전자의무기록은 처방전달시스템에 포함되지 않았던 간호실무 데이터를 실시간 수집하여 간호관찰과 사정, 활동 내역에 대한 추적에 가능하게 한다(Pesut, 2001). 이는 간호실무 데이터의 광범위한 활용을 지원하는 것으로 실무 데이터를 기반으로 한 임상 의사결정 지원 뿐 아니라 임상 연구를 촉진하는 방향으로 발전할 것이다. 본 연구는 이러한 활용 사례로서 실무 욕창발생 예측이라는 문제를 다루었다. 그러나 이러한 접근 가능성과 성공여부는 전자의무기록의 데이터 구조와 질에 의존적이고, 데이터 구조는 활용을 위한 개념 표현(representation)과 추출(extraction)에 의존적이다. 따라서 전자의무기록 시스템의 간호정보 구조에 대한 간호계의 관심이 필요하며 간호 연구자들에게는 데이터 활용과 관련된 간호정보학 지식 및 역량 개발이 요구된다.

본 연구는 연구범위가 단일 의료기관으로 국한되어 구축된 모델의 일반화에 제한이 있을 수 있다. 즉 모델 구축과 평가에 사용된 데이터가 한 의료기관에서 표출되어 해당 의료기관 환자군에 과적합되었을 가능성이 있다. 따라서 다른 의료기관 대상의 반복연구 또는 다중 의료기관 참여 연구가 필요하며, 실무에 직접 적용하기 위해서는 모델에 대한 전향적 재검증 과정과 임상적 효과 근거로서 실무 검증 연구가 필요하다.

결론

중환자실 성인 환자를 대상으로 구축된 BN모델의 정확성과 예측력은 모두 중상위 수준이었으며 동일한 데이터를 이용한 LR모델과 유사한 결과였다. 그러나 실무 사례 커버리지는 BN모델이 LR모델보다 긍정적인 결과를 보여주었다. 따라서 전자의무기록 데이터를 이용한 욕창발생 예측 BN모델 예측력과 실무 적용 가능성은 충분히 수용할 만한 수준이라고 할 수 있다. 이러한 결과는 간호실무에서 BN모델 적용 잠재성을 보여주는 사례로서 고무적이며, 향후 욕창발생 이외 다양한 간호실무의 불확실성을 효과적으로 다루는

새로운 전략이 될 수 있다. 또한 구현을 통한 실무 활용은 지속적인 모니터링이 필요한 환자에게 간호업무 부담을 감소시켜 줄 뿐만 아니라 실수에 의해 또는 무의식적으로 간과될 수 있는 위험 환자를 찾아줌으로써 기존 위험사정 도구를 보완하고 효과적인 예방활동을 촉진하는 데 기여할 것이다.

REFERENCES

- Baumgarten, M., Margolis, D. J., Localio, A. R., Kagan, S. H., Lowe, R. A., Konsian, B., et al. (2008). Extrinsic risk factors for pressure ulcers early in the hospital stay: A nested case-control study. *The Journals of Gerontology Series A: Biological Sciences & Medical Sciences*, 63A, 408-413.
- Beekman, D., Schoonhoven, L., Fletcher, J., Furtado, K., Gunningberg, L., Heyman, H., et al. (2007). EPUAP classification system for pressure ulcers: European reliability study. *Journal of Advanced Nursing*, 60, 682-691. doi:10.1111/j.1365-2648.2007.04474.x
- Bergstrom, N., Braden, B., Kemp, M., Champagne, M., & Ruby, E. (1998). Predicting pressure ulcer risk: A multisite study of the predictive validity of the Braden Scale. *Nursing Research*, 47, 261-269. doi:10.1097/00006199-199809000-00005
- Brown, S. J. (2004). The Braden Scale: A review of the research evidence. *Orthopaedic Nursing*, 23, 30-38. doi:10.1097/00006416-200401000-00010
- Burnside, E. S., Rubin, D. L., & Shachter, R. D. (2004). Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. *Studies in Health Technology And Informatics*, 107(Pt 1), 13-17.
- Cho, I., & Noh, M. (2010). Braden Scale: Evaluation of clinical usefulness in an intensive care unit. *Journal of Advanced Nursing*, 66, 293-302. doi:10.1111/j.1365-2648.2009.05153.x
- Cho, I., Yoon, H. Y., Park, S. I., & Lee, H. S. (2008). Availability of nursing data in an electronic nursing record system for a development of a risk assessment. *Journal of Korean Society of Medical Informatics*, 14, 161-168.
- Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 26, 1-24. doi:10.1016/S0933-3657(02)00049-0
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons.
- De Laat, E. H., Pickkers, P., Schoonhoven, L., Verbeek, A. L., Feuth, T., & van Achterberg, T. (2007). Guideline implementation results in a decrease of pressure ulcer incidence in critically ill patients. *Critical Care Medicine*, 35, 815-820. doi:10.1097/01.CCM.0000257072.10313.56
- Defloor, T., & Grypdonck, M. F. (2005). Pressure ulcers: Validation of two risk assessment scales. *Journal of Clinical Nursing*, 14, 373-382. doi:10.1111/j.1365-2702.2004.01058.x
- Fernandez, G. (2010). *Statistical data mining using SAS applications* (2nd ed.). Minnesota: CRC Press.
- Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine*, 29, 1043-1051. doi:10.1007/s00134-003-1761-8
- Fogerty, M. D., Abumrad, N. N., Nanney, L., Arbogast, P. G., Poulouse, B., & Barbul, A. (2008). Risk factors for pressure ulcers in acute care hospitals. *Wound Repair Regeneration*, 16, 11-18. doi:10.1111/j.1524-475X.2007.00327.x
- Harrison, J. H. (2008). Introduction to the mining of clinical data. *Clinics in Laboratory Medicine*, 28, 1-7. doi:10.1016/j.cll.2007.10.001
- Hatanaka, N., Yamamoto, Y., Ichihara, K., Mastuo, S., Nakamura, Y., Watanabe, M., et al. (2008). A new predictive indicator for development of pressure ulcers in bedridden patients based on common laboratory tests results. *Journal of Clinical Pathology*, 61, 514-518. doi:10.1136/jcp.2007.050195
- Jones, K. R., Fennie, K., & Lenihan, A. (2007). Evidence-based management of chronic wounds. *Advances in Skin & Wound Care*, 20, 591-600. doi:10.1097/01.ASW.0000284936.32707.8d
- Korb, K. B., & Nicholson, A. E. (2004). *Bayesian artificial intelligence*. London: Chapman & Hall/CRC press.
- Lee, K. O., Shin, H. J., Park, H. A., Jeong, H. M., Lee, M. H., Choi, E. H., et al. (2000). Patient severity classification in a medical ICU using APACHE III and patient severity classification tool. *Journal of Korean Academy of Nursing*, 30, 1243-1253.
- Lee, S. M., Abbott, P., & Johantgen, M. (2005). Logistic regression and Bayesian networks to study outcomes using large data sets. *Nursing Research*, 54, 133-138. doi:10.1097/00006199-200503000-00009
- Lee, Y., Jeong, I., & Jeon, S. (2003). A comparative study on the predictive validity among pressure ulcer risk assessment scales. *Journal of the Korean Academy of Nursing*, 33, 162-169.
- Lucas, P. J. F., van der Gaag, L. C., & Abu-Hanna, A. (2004). Bayesian network in biomedicine and health-care. *Artificial Intelligence in Medicine*, 30, 201-214. doi:10.1016/j.artmed.2003.11.001
- Park, H. S., Park, K. Y., & Yu, S. M. (2005). Factors influencing the development of pressure ulcers in surgical patients. *Journal of Korean Academy of Nursing*, 35, 125-134.
- Pesut, D. (2006). 21st century nursing knowledge work: Reasoning into the future. In Weaver, C., Delaney, C., Weber, P. & Carr, R. (Eds.), *Nursing and informatics for the 21st century: Cases, practice, and the future*. (pp. 14-17). Chicago: HIMSS Press.
- Russo, C. A., Steiner, C., & Spector, W. (2008). *Hospitalizations related to pressure ulcers, 2006*. (Rep. No. HCUP Statistical Brief #64). Rockville, MD: Agency for Healthcare Research and Quality.
- Schoonhoven, L., Haalboom, J. R. E., Bousema, M. T., Algra, A., Grobbee, D. E., Grypdonck, M. H., et al. (2002). Prospective cohort study of routine use of risk assessment scales for prediction of pressure ulcers. *British Medical Journal*, 325, 797-800. doi:10.1136/bmj.325.7368.797
- Thede, L. Q., & Sewell, J. P. (2010). *Informatics and nursing: Competencies and applications* (3rd Ed.). Philadelphia: Wolters Kluwer Health | Lippincott Williams & Wilkins.