

의무기록의 다각적 활용을 통한 충실도 높은 병원 암등록 체계의 구축: 서울아산병원의 경험

김화정¹, 조진희^{1,2}, 유용만^{1,3}, 이선혜^{1,2}, 황경하^{1,2}, 이무송^{1,4}

¹서울아산병원 암센터; ²서울아산병원 의무기록팀; ³서울아산병원 의료정보팀; ⁴울산대학교 의과대학 예방의학교실

Construction and Validation of Hospital-Based Cancer Registry Using Various Health Records to Detect Patients with Newly Diagnosed Cancer: Experience at Asan Medical Center

Hwa Jung Kim¹, Jin Hee Cho^{1,2}, Yongman Lyu^{1,3}, Sun Hye Lee^{1,2}, Kyeong Ha Hwang^{1,2}, Moo-Song Lee^{1,4}

¹Asan Cancer Center, Asan Medical Center; ²Medical Record Department, Asan Medical Center; ³Asan Medical Center; ⁴Department of Preventive Medicine, University of Ulsan College of Medicine

Objectives: An accurate estimation of cancer patients is the basis of epidemiological studies and health services. However in Korea, cancer patients visiting out-patient clinics are usually ruled out of such studies and so these studies are suspected of underestimating the cancer patient population. The purpose of this study is to construct a more complete, hospital-based cancer patient registry using multiple sources of medical information.

Methods: We constructed a cancer patient detection algorithm using records from various sources that were obtained from both the in-patients and out-patients seen at Asan Medical Center (AMC) for any reason. The medical data from the potentially incident cancer patients was reviewed four months after first being detected by the algorithm to determine whether these patients actually did or did not have cancer.

Results: Besides the traditional practice of reviewing the charts of in-patients upon their discharge, five more sources of information were added for this algorithm, i.e., pathology reports, the national severe disease registry, the reason for treatment, prescriptions of chemotherapeutic agents and radiation therapy reports. The constructed algorithm was observed to have a PPV of 87.04%. Compared to the results of traditional practice, 36.8% of registry failures were avoided using the AMC algorithm.

Conclusions: To minimize loss in the cancer registry, various data sources should be utilized, and the AMC algorithm can be a successful model for this. Further research will be required in order to apply novel and innovative technology to the electronic medical records system in order to generate new signals from data that has not been previously used.

Key words: Completeness, Hospital-based cancer registries, Medical records linkage, Tumor registration

J Prev Med Public Health 2010;43(3):257-264

서 론

암등록사업은 암환자의 진단 및 치료관련 정보를 체계적으로 수집, 관리하는 건강정보체계의 한 부분으로, 암의 발생률과 암환자의 생존율 등의 암질환 관련 통계를 생성하는 암역학연구의 주요 정보원이다 [1-3]. 우리나라 암등록사업은 수련병원들을 중심으로 하는 중앙암등록사업과 지역내 전 주민을 대상으로 하는 지역암등록사업으로 구성된다 [1,4]. 중앙암등록사업과 지역암등록사업은 각각 1980년과 1983년에 시작되었으나, 전국규모의 암발생 통계를 산출한

것은 2002년 이후의 일이다 [5,6].

보건사업부(현재 보건복지부)의 주관으로 전국 50여 개 수련교육병원을 대상으로 중앙암등록사업이 시작된 이래 참여기관 수가 지속적으로 증가하였으나, 2002년도 사업실적보고서에 따르면 참여대상 병원 중 실제 참여율은 90.2%(대상병원 154개 중 139개 병원에서 중앙암등록사업에 참여)에 그치는 것으로 발표되었다 [1]. 또한 중앙암등록사업 참여대상이 아닌 중소병원에서 암을 진단받고 해당기관에서 지속적으로 치료받는 환자의 경우 암발생 통계에서 누락되므로 전국의 모든 암 발생자가 등록된다고 할 수 없

을 뿐 아니라, 분모에 해당하는 인구집단을 규정할 수 없다. 더욱이 2003년 암관리법이 제정되어 의무화되기 전까지의 무기록사들의 자발적인 참여로 수행되어 온 관행에 따라, 몇몇 대형병원을 제외하고 “조직학적 확진 유무와 관계없이 병원에서 진단 또는 치료한 입원, 외래, 응급 등의 암환자는 모두(암 이외의 질환으로 내원한 환자 포함) 등록한다.”는 지침 [7]과 달리 현실적으로는 대부분 입원환자의 퇴원 당시 의무기록 분석결과를 주된 자료원으로 암환자 여부를 파악하여 등록되었다. 최근에는 대형병원들을 중심으로 등록병원에서 중증진료등록자료 등의 자료원을 활용하여 외래방문 암환자의 등록이 증가되고 있으나, 암환자 등록실패를 전적으로 장담할 수는 없다.

반면, 지역암등록사업의 경우 해당 지역의 암발생률을 산출할 수 있다는 장점을 가지나, 사업 수행 지역이 제주도를 제외하고는 모두 대도시(광주, 대구, 대전, 부산, 서울, 울산, 인천)로 국한되어 있어 [5,6], 국가 암발생 통계를 추정하는데 제한이 있다.

이러한 한계점을 보완하기 위해 중앙암등록본부는 1991년부터 서울지역암등록사업과의 연계를 모색하기도 하였다. 특히 1999년 보건복지부의 암관리사업 강화로 국가암발생통계 생산의 중요성이 증대됨에 따라, 중앙암등록본부에서는 중앙암등록자료 이외에 지역암등록자료, 학술단체에서 운영하는 암종별 암등록자료, 국민건강보험공단 암발생 추정 데이터베이스(의료기관이 건강보험공단에 청구하는 수진자료), 그리고 통계청 암사망자료들을 통합하여 국가암발생 데이터베이스(Korea National Cancer Incidence Database, KNCI DB)를 구축하고, 이를 통하여 국가암발생 통계를 산출하고 있다 [6,8].

그러나 이러한 노력에도 불구하고 KNCI DB보다 지역암등록자료를 활용한 국가암발생률 추정이 상대적으로 더 안정적이라는 보고가 있었다 [9]. 그러나 보건복지부가 지정된 8개 지역암등록본부에서 운영되던 지역암등록사업이 2007년 이후 서울지역암등록사업의 중단으로 축소 운영되면서, 성별, 암종별 과대 혹은 과소 추정의 문제가 발생하고 있다 [5]. 특히 서울은 전 국민의 22%에 해당하는 인구 [10]가 거주하는 수도로서, 서울 지역암등록사업의 중단은 다른 지역암등록본부의 폐쇄보다 큰 타격이 예상되며, 궁극적으로 국가암발생통계에 미칠 영향을 간과할 수가 없다. 비록 중앙암등록사업 참여기관 중 서울지역 수련병원들에서의 병원 암등록사업의 충실도 제고가 근본적 해결방안은 아니라 하더라도, 현 시점에서 우선적으로 수행 가능한 지역암등록사업 보완 방법으로 제시될 수 있겠다.

아울러 중앙암등록이 이전의 등록기관으로부터 등록된

암환자만을 수집, 관리하는 수동적 시스템이 아니라, 암사망자료, 중증진료등록자료, 암수진자료 중 등록되지 않은 환자에 대해 암발생통계조사를 병행하는 능동적 암등록시스템을 구축하였음에도 불구하고, 등록시기가 지연된다는 제한점을 여전히 가진다. 우리나라 중앙암등록사업은 현재 연간 암발생 건수의 약 90%정도를 등록받고 있는데, 진단 및 치료기술의 발전과 조기검진의 보편화로 입원이 아닌 외래진료 기반의 암 진단 및 처치 비중이 증대됨에 따라, 점차적으로 중앙암등록의 완전성이 위협받을 것으로 예상된다.

이에 본 연구는 서울아산병원의 병원의무기록을 다각적으로 활용하여 누락을 최소화하고, 시의적절하게 등록하는 병원암등록 시스템을 구축하고, 그 완전성을 평가해 보고자 한다.

대상 및 방법

1. 잠재적 암환자 추출을 위한 알고리즘의 구축

2007년 1월 1일부터 1월 31일까지 서울아산병원에 내원한 외래, 입원, 응급환자들을 대상으로 잠재적 암환자 (potential cancer patient case)를 추출하는 알고리즘을 개발한다. 기존의 서울아산병원 암등록사업의 대상자는 퇴원 환자 전원 대상의 의무기록 분석과정에서 암에 해당하는 상병을 가진 환자(ICD-10 기준 C00~C97, D00~D09)로 정의되었다. 새로 개발되는 잠재적 암환자 추출을 위한 알고리즘 적용 전과 후의 암등록사업 완결성을 비교하기 위하여, 기존 암등록사업 대상자 탐색 방법에 해당하는 퇴원시점에서 진료기록 분석 결과 해당환자의 주상병 혹은 부상병에 암질환이 포함여부를 알고리즘의 1단계로 설정하였다.

나아가 본 연구의 목적이 병원의무기록을 다각적으로 활용하여 누락을 최소화하는 병원암등록 시스템을 개발하는데 있으므로, 입원 뿐 아니라 외래 및 응급실 방문 시의 각종 진료기록, 암환자가 자주 방문하는 특정과 외래나 클리닉의 방문환자명단, 항암제를 포함한 각종 의약품의 처방 및 투약 기록, 각종 검사, 처치 및 수술의 처방 기록, 혈액검사 보고서를 포함한 진단검사보고서, 영상검사보고서, 병리보고서(조직검사보고서 및 세포검사보고서 포함), 방사선치료보고서, 수술결과보고서, 암등록명단, 중증환자등록명단과 본원에서 발급된 사망진단서, 부검기록, 타원에서 진료받은 기록지 및 진료의뢰서 등의 다양한 정보를 활용하여 기존 암등록 사업에 등록되지 않은 추가 등록대상자를 파악하였다. 즉, 기존의 암등록사업에서 암등

록 대상자이나 압등록이 되지 않는 누락 사례(즉, 실제로 서울아산병원에 내원한 압환자이나, 서울아산병원 압등록 사업에 등록되지 않는 사례)는 대다수 입원 이외의 경로로 내원하는 환자들 중에서 발생할 것으로 예상되므로, 이러한 누락 사례를 최소화하는 방안으로 입원기록과 경과기록, 간호기록 그리고, 퇴원요약지 등 알고리즘의 1단계에서 사용된 정보원 이외의 다양한 정보원을 추가적으로 활용하였다. 효율적인 알고리즘의 구축을 위하여 앞서 기술된 여러 정보원 중 실마리정보(signal)로의 활용 가능성을 평가 후, 선택된 실마리정보 별 양성예측도를 측정하여 해당 실마리정보의 양성예측도가 높은 순서대로 알고리즘을 구축하였다.

기존의 입원환자 위주의 병원압등록과 다각적인 의무기록 활용을 통하여 누락을 최소화하려는 새로운 병원압등록 시스템의 완결성의 변화를 측정하기 위해 대상자들의 병원 방문경로를 파악하였다. 해당기간 중 실마리정보가 1개 이상 발생 시 입원, 외래, 응급실의 순서로 우선하였다. 예를 들어, 동일인이 2007년 1월 3일 외래에서 항암제를 투여받고, 이후 2007년 1월 5일에 부작용으로 응급실을 내원한 후 익일에 입원을 하였을 때, 해당환자의 방문경로를 입원으로 처리하여, 기존의 압등록사업에서 누락되는 환자의 수를 추정하였다.

2. 알고리즘의 타당도 평가

2008년 4월 1일부터 30일까지 서울아산병원에 내원한 외래, 입원, 응급환자를 대상으로 구축된 알고리즘을 적용하여 추출된 잠재적 압환자 3115명 중 500명을 무작위추출하여 알고리즘의 타당도를 평가하였다. 당시 서울아산병원의 압등록 시스템은 중앙압등록에 보고하는 필수 압등록 항목 이외에 각 과별 협의를 통하여 압종별로 특화된 10종류의 압등록 정보요약양식을 개발 예정이었기에, 특화된 압등록 정보요약양식을 보유하는 압종에 해당하는 두경부암(갑상선암 포함), 위암, 소장 및 대장암, 간암, 담낭암, 기관지 및 폐암, 유방암, 부인과암, 비뇨기암(전립선암 포함), 혈액암 환자 각 40명과 특화된 압종에 해당되지 않는 기타암에 해당하는 100명을 층화 무작위추출하여 검증하였다. 암질환의 초진 후 유병기간이 길어질수록 암질환 확진 및 치료 관련 정보가 축적되는 것이 당연하므로, 알고리즘을 통해 추출된 잠재적 압환자들을 대상으로 암질환 확진여부의 판단은 실마리정보가 처음 발견된 이후 검토시점이 늦어질수록 정확할 것이나, 효율적인 압등록 사업의 수행을 위하여 각각의 실마리정보가 발생된 4개월 이후의 시점을 기준으로

하였다. 이는 '첫 진단 후 4개월 이내에 해당 암의 원발부위와 전이부위에 대하여 시행된 치료'만을 입력하는 중앙압등록의 치료관련 정보 수집원칙과 일치되며 [7, 본원의 각 압종별 임상진료의 의견을 반영하여 1차적인 진단과 치료가 마무리되는 시기로 합의된 시점으로 임상적으로 타당하다.

이후 타당도 평가과정에서 확인된 문제점들을 기준으로 알고리즘을 수정·보완하고, 변경된 알고리즘의 타당도 평가를 위하여 2차 타당도 평가를 수행하였다. 두 번째 자료원은 2009년 1월 1일부터 3월 31일까지 서울아산병원에 내원한 외래, 입원, 응급환자 중 수정된 알고리즘을 적용하여 잠재적 압환자로 의심되는 6669명 전원을 대상으로, 첫 번째 타당도 평가과정과 동일하게 알고리즘의 타당도를 평가하였다. 1차 타당도 평가 시기 이후 전술한 10종의 특화된 압등록 정보요약양식 이외에도 갑상선암과 골연부육종, 소아암 대상의 특화된 압등록 정보요약양식이 추가되어 두경부암, 갑상선암, 위암, 소장 및 대장암, 간암, 담낭암, 기관지 및 폐암, 유방암, 부인과암, 비뇨기암(전립선암 포함), 혈액암, 골연부육종, 소아암, 그리고 기타암으로 분류하여 검증하였다.

결 과

1. 잠재적 압환자 추출을 위한 알고리즘의 구축

우선 알고리즘의 1단계로 퇴원환자의 의무기록 분석을 수행하였다. 2007년 1월 1일부터 1월 31일까지 서울아산병원에 내원한 입원환자 중, 퇴원시점을 기준으로 해당 환자의 상병 목록에 암질환이 포함되는 1665명을 대상으로 암질환 확진여부를 확인하여 1169명을 압등록하였다.

앞서 기술된 정보원 중 코드화 되어있지 않고, 텍스트로 기술된 정보로 구성된 정보원은 활용이 어려워 제외시켰다. 그 결과 각종 처방(의약품, 검사, 처치, 수술 등 포함) 시 상병명, 병리검사결과, 진단검사결과(혈액암일 경우), 항암제 처방 및 투약자 명단, 치료방사선과 내원환자 명단, 중증환자등록명단 중 압환자명단 등 5가지 정보원을 2단계 실마리정보로 활용할 수 있음을 파악하였다. 영상검사보고서의 경우 해당 정보의 요약 및 분류가 어렵고, 영상검사만으로 암진단이 가능한 간암과 조직을 확보하기 어려운 담낭암의 경우 각종 처방 시 상병명, 항암제 처방 및 투약자 명단, 치료방사선과 내원 환자 명단에 포함되어 중복되므로 활용하지 않았다. 다만 알고리즘을 적용하여 추정된 잠재적 압환자의 암의 확진 여부의 판별을 위해서는 활용하였

Table 1. Construction of AMC algorithm designed to detect potentially incident cancer patients in order of high positive predictive values (PPVs) for each potential case selection criteria from various sources of information at Asan Medical Center (AMC) during January 2007 (n=1665)

Step	Source of information	n	PPV of each step (%)	Cumulative PPV (%)
# 1	Chart review of in-patient at discharge by well trained RRA [†]	1152	99.74	99.74
# 2a	Radiation therapy report	10 [‡]	100.00 [‡]	98.88
# 2b	Order for chemotherapy	48 [‡]	97.92 [‡]	95.04
# 2c	Pathology report	74 [‡]	97.30 [‡]	89.72
# 2d	Reason for treatment	418 [‡]	95.45 [‡]	70.34
# 2e	National severe disease registry	48 [‡]	79.17 [‡]	70.21

[†] RRA, registered medical record administrator, [‡] Number of patients who meets each criterion among patients with the exclusion of step 1.

[‡] PPV of each criterion among patients who were ruled out from step 1.

Table 2. Characteristics of potential and confirmed cancer patients detected by traditional method and AMC algorithm designed to detect potentially incident cancer patients at Asan Medical Center (AMC) during January 2007 (n=1665)

Demographics	Potential case (n=1152)	Confirmed case (n=1149)
1) Of patients by traditional method		
Age (y)		
Mean ± SD	55.0 ± 13.8	55.0 ± 13.8
Less than 30	33 (2.9 %)	33 (2.9 %)
30 - 39	108 (9.4 %)	106 (9.3 %)
40 - 49	273 (23.7 %)	273 (23.8 %)
50 - 59	288 (25.0 %)	287 (25.0 %)
60 or older	450 (39.1 %)	450 (39.2 %)
Sex		
Male	609 (52.9 %)	609 (53.0 %)
Female	543 (47.1 %)	540 (47.0 %)
2) Of patients by constructed algorithm (N=1665)		
Age (%) (y)		
Mean ± SD	54.0 ± 15.2	54.9 ± 13.9
Less than 30	92 (5.5)	36 (3.1)
30 - 39	153 (9.2)	107 (9.2)
40 - 49	370 (22.2)	274 (23.4)
50 - 59	413 (24.8)	297 (25.4)
60 or older	637 (38.3)	455 (38.9)
Sex		
Male	867 (52.1 %)	617 (52.8 %)
Female	798 (47.9 %)	552 (47.2 %)
Visit route*		
Admission	1228 (73.8 %)	1153 (98.6 %)
OPD (outpatient department)	409 (24.6 %)	16 (1.4 %)
Emergency room	28 (1.7 %)	0 (0.0 %)

* In case of multiple signal generation, signal generated at admission was prior to signal from OPD or emergency room visit, as well as signal generated at OPD to signal from emergency room visit.

다. 수술기록 및 수술결과보고서의 경우 수술과 관련된 상병이 반드시 포함되므로, 처방 시 상병명과 일치하여 중복된 자료원으로 파악되었다. 그러나, 수술기록 및 수술결과보고서의 경우 영상검사보고서의 경우와 마찬가지로 암의 확진 여부 판별 근거로 사용되었다. 사망진단서와 부검보고서의 경우 또한 코드화되어 있지 않을 뿐 아니라, 도착 시 사망(death on arrival, DOA)의 경우를 제외하고 병원이용에 따른 각종 기록들이 발생되므로 잠재적 암환자 추출에

적극적으로 활용되지 않았으나, 잠재적 암환자의 암등록여부를 결정하는 데는 활용하였다. 진료의뢰서나 타원 진료 기록지의 경우 스캔 후 광파일의 형태로 의무기록에 포함되어 진단인력이 진료의뢰서를 모두 검토하지 않는 한 정보원으로 활용하기 어려울 뿐 아니라, 암질환 확진 후 본원으로 전원이 되었다고 하더라도 타원에서 확진 시 활용된 해당조직 검체를 재판독하도록 규정되어 있어 앞서 기술된 6가지 정보원에서 반드시 1개 이상의 실마리정보가 생성되게 되어있어, 이것으로 대체하였다. 다만 영상검사보고서와 마찬가지로 잠재적 암환자의 암등록여부를 결정하는 데는 활용하였다. 요약하자면 해당기간 퇴원환자 전원에 대하여 입원기간 중 상병에 암질환이 포함되는 경우 이외에도 다음에 기술된 5가지 조건에 해당 시 각각의 실마리정보가 발생하여, 총 6가지 중 하나 이상의 실마리정보 발생시 알고리즘에 의해 추정된 잠재적 암환자로 간주하며, 입원환자의 의무기록 분석 이외에 알고리즘에 활용할 실마리정보원은 다음과 같다.

- 1) 해당기간 중 입원, 외래, 응급환자에게 투약, 처치, 수술 등 처방 시 상병에 암질환이 포함되는 경우
- 2) 해당기간 중 병리검사나 진단검사 결과 암으로 판정된 경우
- 3) 해당기간 중 입원, 외래, 응급환자의 투약 중 항암제 코드
- 4) 해당기간 중 방사선 종양학과에서 방사선 치료를 받은 경우
- 5) 해당기간 중 중증질환 등록 명단 중 등록상병이 암질환인 경우

기술된 다양한 정보원을 활용하여 추출된 잠재적 암환자 중 이미 원내 암등록사업에 등록된 환자들을 제외하고, 접근 가능한 환자의 모든 원내 기록을 분석하여 해당 환자의 암질환 확진 사실이 명시된 경우를 병원 암등록 대상으로 정의한다. 즉, 본원을 방문한 이유가 암과 관련이 없는 질병

Table 3. Positive predictive values (PPVs) of AMC algorithm designed to detect potentially incident cancer patients among randomly selected sample patients who visited Asan Medical Center (AMC) during April 2008 (n=500)

Step	Source of information	n	PPV of each step (%)	Cumulative PPV (%)
# 1	Chart review of in-patient at discharge by well trained RRA*	214	99.54	99.54
# 2	Radiation therapy report	54	92.60	97.97
		32 [†]	87.50 [‡]	
# 3	Order for chemotherapy	128	75.79	88.32
		62 [†]	50.00 [‡]	
# 4	Pathology report	169	99.41	88.86
		24 [†]	95.84 [‡]	
# 5	Reason for treatment	433	87.30	82.40
		168 [†]	69.65 [‡]	
# 6	National severe disease registry	131	99.24	82.40
		0 [†]	(-) [‡]	

* RRA, registered medical record administrator.

[†] Number of patients who meets each criterion among patients with the exclusion of previous step.

[‡] PPV of each criterion among patients who were ruled out from previous step.

이라 하더라도, 타 기관에서 암으로 진단 받은 과거력이 있으며 현재 암이 완치되지 않았다고 판단되면 등록하도록 하였다. 그 결과, 2007년 1월 1일부터 1월 31일까지 1달의 기간 동안 서울아산병원에 내원한 잠재적 암환자 1665명을 대상으로 초진 후 4개월이 경과하기 전에 병원암등록사업에 등록되었는지 여부를 기준으로 양성예측도를 각 실마리 정보 별로 계산하여, 양성예측도가 높은 순서대로 알고리즘에 포함시켜 구축하였다 (Table 1). 잠재적 암환자들의 일반적 특성은 Table 2와 같다. 남성이 52.1%, 평균연령은 54.0세로 50세 이상이 63.1%를 차지하였다. 입원 이외의 경로로 확인된 잠재적 암환자는 전체의 73.8%에 해당하였다. 실제로 알고리즘 적용 후 암등록된 환자는 1169명으로 기존의 퇴원환자의 의무기록 분석으로 등록된 암환자(1149명) 이외의 암환자를 추가적으로 발견하는 것으로 확인되었다.

2. 알고리즘의 타당도 평가

첫 번째 자료원은 2008년 4월 1일부터 30일까지 서울아산병원에 내원한 암등록대상자 중 무작위추출한 500명이 대상으로 실마리정보가 발생한 시점에서 4개월이 경과하는 기간 동안의 의무기록을 확인한 알고리즘의 양성예측도는 82.4%였다. 즉, 알고리즘 적용 결과 잠재적 암환자일 것으로 생각되었으나, 실제로 암환자가 아닌 경우가 88명인 것으로 관찰되었다 (Table 3).

알고리즘 상 잠재적 암환자로 분류되지만, 실제로 암질환이 확진되지 않은 대상들에서 거짓 실마리정보가 발생하는 주된 이유는 처방상병과 항암제 처방코드와 관련된 것으로 확인되었다. 암질환을 이유로 각종 처방을 하였어도 암등록되지 않은 환자들은, 추가적인 검사나 종합적 판단

으로 궁극적으로는 암이 아닌 것으로 확인 된 경우(39건)나 임상적으로 혹은 영상검사 결과로 암질환이 진단되었고, 암치료를 수행하였으나 조직을 확보할 수 없는 경우, 또는 노령 등의 이유로 병리검사를 통한 확진이 안된 경우(6건), 아직 진단되지 않은 암질환 의증 환자인 경우(3건) 등이 있었다 (Table 4). 또한 6가지 실마리정보 중 항암제 시그널만 양성인 경우 중 진료과가 안과인 경우, 당뇨병의 합병증인 망막병증(retinopathy)를 치료하기 위하여 항암제를 사용하는 사례였다. 마찬가지로 항암제 시그널만 양성인 포상기태(hydatid mole) 등의 산부인과 환자는 진단명이 암에 해당하지 않으나, 항암제로 치료하는 경우가 빈번함을 확인할 수 있었다. 따라서, 해당과 임상 의에게 문의하여 암환자가 아니나 항암제로 치료하는 진단명(D180, D25, D250, D251, D252, D259, D27, D761, E228, E230, E258, E282, E301, E343, E348, E45, H340, H341, H342, H343, H344, H345, H346, H347, H428, N830, N831, N832, N858, N97, N970, N971, N972, N973, N974, N978, N979, Q771, Q774)의 경우 항암제 이외의 시그널이 모두 음성일 때 암등록 대상자로 추출되지 않도록 조절하였다. 이렇게 알고리즘을 수정한 결과 양성예측도는 기존의 82.4% (Table 3)에서 86.4%로 개선되었음을 확인할 수 있었다.

또한 연구대상에 따라 양성예측도가 높은 순서의 실마리 정보가 상이하고(Table 1,3), 본 연구가 효율적인 암환자 추출을 위한 알고리즘이 아니라, 가급적 누락이 적은 알고리즘의 구축이라는 목적아래 진행됨에 따라, 활용가능한 6가지 실마리정보 중 양성예측도 기준으로 실마리정보의 정보원을 선택하는 것이 아니라, 모두 활용하되 업무의 효율성을 위하여 실마리정보 활용순서를 ① 퇴원환자의 의무기록 분석 결과, ② 병리검사결과보고서 및 진단검사결과보고서, ③ 중증질환등록, ④ 처방 시 상병, ⑤ 항암제 투약, 그

Table 4. Reason for the incorrect identification of cancer patients according to initial AMC algorithm among randomly selected sample patients who visited Asan Medical Center (AMC) during April 2008 (N=500)

Source of information	Reason for false positive signal	Number (%)
Chart review of in-patient	· Confirmed as non-cancer afterwards	1 (1.1)
National severe disease registry	· Cancer diagnosis without pathologic confirm	1 (1.1)
Pathology report	· Confirmed as non-cancer afterwards	1 (1.1)
Radiation therapy report	· Confirmed as non-cancer afterwards (1)	4 (4.5)
	· Confirmed as cancer after the validation period (4 month since diagnosis) (2)	
	· RTx for non-cancer disease (1)	
	- such as, scleroderma	
Reason for treatment	· Confirmed as non-cancer afterwards (35)	51 (58.0)
	· Cancer diagnosis without pathologic confirm (5)	
	· Follow-up loss before additional work-up (3)	
	· Others (8)	
Order for chemotherapy	· Confirmed as non-cancer (1)	30 (34.1)
	· CTx for non-cancer disease (28)	
	- such as, retinopathy, lupus nephritis etc	
	· Others (1)	
Total		88 (100.0)

Table 5. Positive predictive values (PPVs) of AMC algorithm designed to detect potentially incident cancer patients among randomly selected sample patients who visited Asan Medical Center (AMC) from January 2009 to March 2009 (n=6669)

Step	Source of information	n	PPV of each step (%)	Cumulative PPV (%)
# 1	Chart review of in-patient at discharge by well trained RRA*	3710	98.95	98.95
# 2	Pathology report	3858	98.03	97.70
		994 [†]	93.06 [‡]	
# 3	National severe disease registry	2798	97.50	96.83
		250 [‡]	80.40 [‡]	
# 4	Reason for treatment	6030	89.85	89.12
		1556 [‡]	64.59 [‡]	
# 5	Order for chemotherapy	1415	89.26	87.26
		140 [‡]	0.71 [‡]	
# 6	Radiation therapy report	468	95.73	87.04
		19 [‡]	10.53 [‡]	

* RRA, registered medical record administrator.

[†] Number of patients who meets each criterion among patients with the exclusion of previous step.

[‡] PPV of each criterion among patients who were ruled out from previous step.

리고 ⑥ 방사선 종양학과에서의 방사선 치료로 수정하였다. 다만, 기존의 암등록 대상자 탐색 작업이 ①에 의해 이루어졌었던 관계로 기존의 작업에 비하여 새로이 구축된 알고리즘을 활용하는 경우 추가적으로 등록되는 암환자의 규모를 비교하기 위하여 ①이 ②~⑥보다 우선되었다. 이렇게 1차 타당도 검사결과 조정된 알고리즘은 병원암등록 시스템에 도입하여, 2009년 1월 1일 이후 병원 내원 환자들을 대상으로 적용되고 있다.

두 번째 타당도 검사 대상인 2009년 1월 1일부터 3월 31일까지 3달간 서울아산병원에 내원한 환자 중 수정·보완한 알고리즘을 적용 시 잠재적 암환자로 의심되는 6669명 전원에게 실마리정보가 발생한 시점으로부터 4개월 시점까지의 의무기록을 확인하여 알고리즘의 타당도를 평가하였다. 암환자로 의심되는 대상자 6669명 중 해당 기간 내에 확

진을 받은 것으로 확인된 환자는 5805명으로 확인되어 알고리즘의 양성예측도는 87.0%였다. 기존의 퇴원환자의 의무기록 분석으로 등록된 암환자 3671명인데 비하여, 알고리즘으로 추출된 대상자 중 암질환이 확진된 환자는 5805명인 것으로 조사되어, 알고리즘을 적용하지 않았다면 누락되었을 해당기간 암등록 대상자 2958명 중 72.1% (2134명)이 추가적으로 등록되어, 결과적으로 총 암등록환자의 36.8%를 추가적으로 등록되었음을 확인하였다 (Table 5). 최종 알고리즘의 암종별 양성예측도는 위암(ICD10 코드상 C16), 폐암(C34), 대장암(C18-20), 간암(C22), 갑상샘암(C73), 그리고 유방암(C50)에서 각각 96.4%, 94.1%, 95.4%, 91.8%, 98.0%, 98.0%인 것으로 확인되었다.

알고리즘에 의해 생성된 시그널 정보의 개수가 증가함에 따라 암 확진 가능성이 유의하게 증가함을 확인하였고

Table 6. Positive predictive values (PPVs) for number of signal generation according to AMC algorithm designed to detect potentially incident cancer patients among randomly selected sample patients who visited Asan Medical Center (AMC) from January 2009 to March 2009 (n=6669)

Number of satisfied conditions	n	PPV (%)	Test for trend
1	2055	62.24	<0.01
2	894	92.17	
3	1207	98.92	
4	1821	99.73	
5	598	100.00	
6	94	100.00	

($p < 0.01$), 특히 시그널 정보의 개수가 2개 이상일 경우 92.2%의 가능성을 가지고 암환자일 것을 추정할 수 있음이 확인되었다 (Table 6).

고찰

완전하고 타당한 암등록 사업을 수행하기 위해서는 실제 발생한 암환자를 최대한 발견하여 등록함으로써 누락을 최소화하여야 한다 [11]. 본 연구는 병원의료정보 시스템의 다양한 정보원을 다각적으로 활용하여 병원암등록사업의 완전성을 높이고자 수행되었다.

우선 누락을 최소화하는 알고리즘을 구축하기 위하여 이용 가능한 정보원을 효율성과 적절성을 기준으로 분류하였다. 알고리즘의 완결성 제고를 위하여 동일내용을 포함하는 중복된 자료를 제외하고 전부 활용함을 원칙으로 하였으나, 정보가 코드화되어 있지 않고 텍스트의 형태로 서술되어있는 경우는 이용이 어려워 본 연구에서는 포함시키지 못하였다. 그러나, 현재 구축된 알고리즘에서 사용하고 있는 퇴원환자의 의무기록 이외에 병리검사 및 진단검사 결과보고서, 중증질환등록명단, 처방 시 상병정보, 항암제 투약정보, 그리고 방사선 치료명단의 5가지 정보원을 추가하는 것만으로도 암환자의 누락을 10%~40% 정도 피할 수 있는 것으로 확인되어 연구의 목적을 달성하였다고 할 수 있겠다. 건강검진자의 경우에도 암이 의심되는 대상자에서 수행되는 검사과정을 통해 병리검사보고서 및 진단검사결과보고서가 발생되므로 별도의 재확인 과정 없이 통합적으로 활용할 수 있었다.

또한 시그널 발생 후 어느 기간 동안의 의무기록을 분석하여 암확진 여부를 결정할 것인가에 대한 고민도 필요하다. 잠재적 암환자의 의무기록 검토 시기가 늦어질수록 암등록정보의 타당도를 획득할 수 있겠으나, 시기 적절한 통

계산출은 불가능해질 것이다. 한국중앙암등록본부의 암등록지침서의 경우 첫 진단 후 4개월 이내에 시행된 초치료를 입력하도록 명시되어 있어 [7], 본 연구에서도 유사하게 실마리 발생시점 후 4개월의 기간 동안의 의무기록을 검토하여 암환자 여부를 결정하였다. 그러나, 갑상선암의 경우 진단 후 4개월 이내에 어떠한 치료도 받지 않아 병리검사 결과보고서를 통해 실마리정보가 생성된 후 추가적인 어떠한 정보도 발생하지 않은 경우도 많았다. 이와 같이 비치명적 (non-fatal) 암종의 경우 발견 즉시 치료가 진행되지 않고 치료대기시간이 길어져, 해당기간의 의무기록을 검토 시 위양성으로 판명될 수 있어, 이에 대한 고려가 필요하겠다. 암종별 치료대기시간에 맞추어 암 확진여부 판단을 위한 잠재기간(window period), 또는 초진일 기준을 변경하는 방법이 가능하겠다.

반면 새로운 알고리즘을 통한 추가적 등록률을 본문에서 전술된 방법으로 계산하는 것은 충분한 시간이 흐른 후 입원과정을 거치면서 등록될 대상을 빨리 등록할 뿐, 실질적으로 누락될 대상을 찾아낸 것이 아니므로 과대추정되었을 가능성에 대한 우려가 제기될 수 있겠다. 이에 아산병원에서 새로운 암환자 탐색 알고리즘을 적용한 이후 시점인 2009년 1월 1일에서 3월까지 31일까지 내원한 잠재적 암환자 6669명 중 초진 후 8개월의 기간이 경과한 후에 기존 시스템 대비 늘어난 등록례수를 평가하여 4개월의 기준 시점의 결과와 비교하였다. 그 결과 새로운 알고리즘을 통하여 초진일로부터 4개월의 시점까지 추가적으로 등록된 암환자 2958 명 중 초진일로부터 4개월 이후 8개월 이전의 기간 중에 퇴원환자의 의무기록분석 결과 등록되었을 환자는 451 명(15.2%)에 불과한 것으로 파악되었다. 즉, 새로운 알고리즘의 적용으로 기존 시스템에 비해 초기에 등록할 가능성을 배제한 추가 등록률은 27.9% (95% CI = 26.7% - 29.1%)로 나타나, 판단 기준기간을 충분히 고려한 후에도 약 20% 정도의 누락례를 찾아낼 것을 기대할 수 있겠다.

본 연구에서는 알고리즘의 구축과 2차례의 타당도 검증을 위하여 각각의 다른 시기에 내원한 환자들을 대상으로 하였는데, 적용시기와 무관하게 알고리즘의 양성예측도는 80% 이상으로 유지됨이 확인되었다. 그러나, 정보원별 암환자 발견율은 매우 상이하였는데, 위양성인 환자들에 대한 원인들을 추가적으로 파악하여, 알고리즘을 보다 정교하게 만드는 작업이 필요하겠다. 뿐만 아니라, 현재 실마리 정보 생성이전 이외의 추가적인 정보원의 개발도 필요할 것이다. 예를 들어, 영상검사결과보고서를 대상으로 단어 추출 방식 등을 활용하여 코딩하는 방법을 적용해보았으나, 원발부위나 병기 등 암질환 관련 특성에 따라 검색해야

할 표제가 너무나 다양하고, 오류율도 높아 적용하지 못하였다. 향후 원내 차세대 전자의무기록(electronic medical record, EMR) 시스템의 구축으로 이러한 현실은 개선될 것으로 기대되나, 차세대 전자의무기록 시스템의 구현과, 새로운 시스템의 적용 후 적응하기 전까지의 과정에서의 정보수집과 후향적 정보수집을 위해서 암종별로 특화된 표제어를 선별하고, 선별된 표제어들을 안정적으로 검색해내는 기술을 구축된 알고리즘에 적용시키려는 노력이 필요할 것이다.

본 연구의 목적이 완결성 높은 잠재적 암환자 추출 알고리즘의 구현이나, 자료의 특성상 민감도와 특이도를 구할 수 없었다. 등록의 완전성을 파악하는 방법으로 DCO index 나 사망률/발생률 비 등의 지표를 산출하거나, 비슷한 인구구조를 가지는 집단에서의 발생률과의 비교하는 방법이 있겠다 [12]. 그러나, 이러한 방법들은 병원암등록에 적용하기 부적절하고 병원에 내원하는 환자의 특성에 의해 좌우되므로 비슷한 인구구조를 가지는 집단이 아닌 경우, 그 결과를 암등록사업에서 추출된 통계와 비교하기 어렵다. 대체적인 방법으로 연령별 암발생률 곡선의 변화를 관찰하는 방법이 사용되기도 하나, 자료가 축적되어야 의미있는 결과를 도출할 수 있을 것이므로 [13], 향후 수행하고자 한다. 또한 학회차원에서 수행되고 있는 학술목적의 암등록 등 외부자료를 이용하여 capture-recapture 방법을 이용한 자료의 완전성 추정도 가능하겠다.

참고문헌

- Ahn YO. Cancer registration in Korea: the present and furtherance. *J Prev Med Public Health* 2007; 40(4): 265-272. (Korean)
- Greene FL, Gilkerson S, Tedder P, Smith K. The role of the hospital registry in achieving outcome benchmarks in cancer care. *J Surg Oncol* 2009; 99(8): 497-499.
- Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol* 2008; 13(2): 102-111.
- Jin DG, Kam S, Chun BY, Ahn SK, Kim JY. The efficient methods of population-based cancer registration in Daegu city. *Korean J Prev Med* 2002; 35(4): 322-330. (Korean)
- The Community of Population-based Regional Cancer Registries in Korea. A validation of estimating the national cancer incidence in Korea using the databases of 7 population-based regional cancer registries except Seoul. *J Prev Med Public Health* 2009; 42(2): 130-134. (Korean)
- Shin HR. Global activity of cancer registries and cancer control and cancer incidence statistics in Korea. *J Prev Med Public Health* 2008; 41(2): 84-91. (Korean)
- Korea Central Cancer Registry. *Manual for Cancer Registration*. Goyang: National Cancer Center; 2007. p. 51. (Korean)
- Shin HR, Won YJ, Jung KW, Kong HJ, Yim SH, Lee JK, et al. Nationwide cancer incidence in Korea, 1999-2001; first result using the national cancer incidence database. *Cancer Res Treat* 2005; 37(6): 325-331. (Korean)
- The Community of Population-based Regional Cancer Registries in Korea. An estimation of the national cancer incidence in Korea for 2000-2002 using the databases of 8 population-based regional cancer registries. *J Prev Med Public Health* 2008; 41(6): 380-386. (Korean)
- Korea National Statistical Office. *Summary of Census Population by Administrative District in Korea, 2005-2008*. Daejeon: Korea National Statistical Office; 2009. (Korean)
- Parkin DM, Muir CS. Comparability and quality of data. In: Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J. *Cancer Incidence in Five, Volume VI* (IARC Scientific Publications No.120). Lyon: IARC; 1993. p. 45-55.
- Skeet R. Quality and quality control. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. *Cancer Registration: Principles and Methods* (IARC Scientific Publications No.95). Lyon: IARC; 1991. p. 101-107.
- Jensen O, Strom H. Purposes and uses of cancer registration. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. *Cancer Registration: Principles and Methods* (IARC Scientific Publications No.95). Lyon: IARC; 1991, p. 7-21.