

대구지역 암등록사업의 효율적 수행방안

진대구, 감 신, 천병렬, 안순기, 김종연

경북대학교 의과대학 예방의학교실

The Efficient Methods of Population-based Cancer Registration in Daegu City

Dae Gu Jin, Sin Kam, Byung Yeol Chun, Soon Ki Ahn, Jong Yeon Kim

Department of Preventive Medicine and Public Health, School of Medicine, Kyungpook National University

Objective: This study was conducted to automatically improve the completeness and validity of the Daegu Cancer Registry, using cross record linkage of many data sources, and to develop a computerized patient enrollment system for efficient communication among cancer researchers via the internet.

Method: We analyzed 10,229 cancer patients who were reported in the National Cancer Registry, and from pathological reports, health insurance cancer claims lists, cancer patient records at hospital information centers and death certificates from the Korea National Statistical Office.

Result: We confirmed 4,624 cancer patients and found 897 of new cases from a review of medical chart. The new cases were

detected efficiently using cross record linkage. We developed a computerized patient enrollment system, based on a client-server model, for the input of cancer patients, and then developed a web-based reporting homepage and patient enrollment system for the internet.

Conclusion: This system could manage cancer databases systematically, and could be given to other researchers as a basic database.

Korean J Prev Med 2002;35(4):322-330

Key Words: Registries, Medical record linkage, Automated cancer registration

서 론

암은 평균수명의 연장으로 인한 노령 인구의 증가와 생활양식, 식습관, 환경의 변화에 의한 질병구조의 변화 그리고 진단기술의 발달 등으로 점차 발생률 및 사망률이 증가하고 있다. 우리나라에서 암 사망률은 1990년 인구 10만명당 110.6명에서, 2000년 124.2명으로 계속 증가하여 전체 사망원인의 23.8%를 차지하면서 사망순위 1위가 되었다 [1]. 그러나, 최근 선진국에서는 적극적인 관리노력으로 암 발생률 및 사망률이 감소하고 있다. 실제 미국의 경우 국가차원의 지속적인 암 연구와 20년 이상의 암 예방 및 관리의 노력으로 1990년부터는 전반적인 암 발생 및 사망의 감소를 보이고 있다 [1,2].

암사망률의 감소는 암발생률의 감소와 효율적인 암환자 관리가 병행되어야 가능하다. 이를 위한 기초적인 작업이 암등록사업이다. 즉 국가적인 차원에서 암환자를 등록하여 암의 발생률, 유병률 및 사

망률에 대한 정확한 통계를 제공함으로써 암관리사업을 수행하고 나아가 암의 발생원인 연구, 암과 관련된 보건정책 기획 및 평가의 기초자료로서 유용하게 사용될 수 있다.

현재 우리 나라에서 수행하고 있는 암등록사업은 수련병원들을 중심으로 하는 중앙암등록사업(hospital-based national cancer registry)과 지역 내에 사는 전 주민을 대상으로 하는 지역암등록사업(population-based cancer registry)으로 나눌 수 있다.

중앙암등록사업은 1980년 7월 보건복지부 주관으로 전국 47개 레지던트 수련병원이 참가함으로써 시작되어, 1990년 6월 15일부터는 인턴 수련병원까지 확대 실시되었다. 1991년부터는 중앙암등록사업이 병원중심의 등록자료에 기인함으로써 생기는 한계점을 극복하고자 서울지역 암등록사업과 연계를 모색하였다. 하지만, 그 완전성(completeness)면에서 일부 중소병원에서 진단받은 환자가 누락되고

있고, 전국의 자료를 수집하는 과정에서 자료의 타당성 역시 떨어지고 있다. 게다가, 병원중심의 암등록사업 특성 때문에 모집단이 불분명하고, 암등록환자에 대한 추적조사가 이루어지고 있지 않아 국가 암관리에 필요한 주요 암통계치를 적극적으로 산출하지 못하고 있다.

한편, 국내에서는 최초로 1982년부터 경기도 강화군에서 지역암등록사업이 시행되었다. 그러나 강화군은 인구수가 약 80,000명인 농촌지역으로 우리나라를 대표하는 암통계를 생산하기에는 인구수가 작으며 원발부위별로 안정된 통계를 생산하기 위하여서는 장기간의 추적이 필요하다. 본격적인 지역암등록사업은 1990년 서울시 지역암등록추진연구부터 시작되었으며, 그후 보건복지부 암정복추진연구개발사업이 시작되면서 1997년부터는 5대 대도시(서울 1990, 부산 1995, 대구 1997, 광주 1997, 인천 1998) 지역암등록사업으로 확장되어 2002년 현재는 7개 대도시(서울, 부산, 대구, 광주, 인천, 대전, 울산)와 제주에서 시행되고 있다. 그러나 아직도

지역암등록에서 누락된 환자의 비율이 10-20%정도를 차지하고 있으며, 지역별로 누락률과 정확도가 떨어져 이에 대한 보완이 시급한 실정이었다 [3]. 2001년 국립암센터에서 부산지역을 대상으로 1999년에 발생한 암환자 충실도 조사를 시행한 결과 누락률은 4.65%(515건)으로 누락률이 호전되었지만, 여전히 지역 등록소마다 차이가 나고 있다 [4].

이에 본 연구는, 수동적인 등록의 비용 문제와 증가하고 있는 전산화된 임상자료들 그리고, 빈번한 자료교환(transaction)과 자료들간의 표준화요구라는 상황에 부응하여 수동(manual)으로 등록하는 방법과는 달리, 자동화된 암등록(automated cancer registration)을 시도하고, 그 가능성을 평가해 보고자 한다.

우선, 기본작업으로 자료원의 구조를 분석하고 동일한 자료축약(consolidation) 방법을 일관성 있게 적용하여 자료구조를 표준화한 후, 자료원들 간의 연결(cross record linkage)을 시도하여 의무기록을 열람할 때 수동으로 찾을 항목을 줄이고자 노력하였다.

각각의 자료원들에 대해 의무기록열람을 시행하여 암을 확인하고 최종 암등록 자료를 데이터베이스로 저장하였다. 또한, 암등록 전산시스템을 개발하여 암등록을 표준화하고 결과공유를 통해 기초암연구자료로 활용하고자 하였다.

최종적으로, 가능성을 평가하기 위해 최종암등록자료를 다시 병원전산실을 통해 검색한 암환자명단과 연결을 시도하여 의무기록열람을 하지 않고 자동으로 암등록이 가능한 정도를 파악하였다.

하지 않지만 암진단과 치료가 가능하고 자료수집에 적극적으로 협조해주는 19개의 병원 그리고 중앙암등록 참여 병원의 해부병리과 6개와 개원하고 있는 7개의 해부병리과 의원을 대상으로 1999년 암 발생환자의 자료를 수집하였다 [5]. 그리고 대구 이외의 지역에서 암등록이 이루어진 경우는 국립암센터를 통해서 자료를 수집하였다.

자료원은 입원환자를 대상으로 하여 등록 병원이 보고하는 중앙 암등록 자료, 해부병리 결과보고서, 의료보험공단 암수진자 명단, 병원전산실에서 ICD-10 코드 C00-C97로 검색한 암환자 명단, 사망신고서를 토대로 통계청에서 집계한 암에 의한 사망자 명단이다 (Table 1).

2. 연구방법

기본작업으로 자료구조를 표준화한 후, 자료원들 간의 연결을 시도하였다. 그리고, web-based reporting을 통해 암등록을 표준화하며 그 결과를 공유함으로써 기초암 연구자료로 활용하고자 하였다. 최종적으로 자동화된 암등록의 가능성을 평가하였다.

1) 자료원 사전처리

암등록 모식도는 Figure 1과 같고, 자료원별 환례수는 해부병리보고서(A)가 598명, 의료보험공단 수진자 명단(B)은 108,813명, 병원 전산실 암환자 명단은 1,729명(C)이었다 (Figure 2).

우선, 중앙암등록에 참여하는 6개 종합병원의 해부병리과와 7개 해부병리과 의원에 해부병리보고서(병원 의무기록번호,

환자명, 진단명, 진단코드, 형태코드)를 요청하였다. 자체 등록프로그램을 쓰는 3곳의 대학병원 해부병리과에는 직접 등록 프로그램을 검색하여 암환자명단을 얻었는데 암등록 여부와 대구 거주 확인을 위해서는 주민등록번호와 주소지 정보가 필요하기에 미비한 항목에 대해 병원 의무기록번호로 병원 전산실 암환자명단과 연결하였다. 연결이 되지 않는 경우만 병원전산실에 요청하였다. 결과보고시스템이 병원정보시스템에 통합되어 있는 2곳의 해부병리과는 전산팀의 협조를 얻어 명단을 입수했다. 나머지 병원은 직접 결과보고서를 열람하여 입력하였다.

그리고, 중앙암등록사업 본부로부터 대구시내 소재한 병원에서 1999년 의료보험공단에 암환자로 청구한 수진자명단을 입수하여, 해당 병원의 의무기록번호, 환자주소를 알아내기 위해 주민등록번호로 병원 전산실에 의뢰하였다.

이런 후에 해부병리과를 거치지 않고 암으로 진단 받은 환자를 추가적으로 찾기 위해서는 임상병리과, 진단방사선과, 치료방사선과의 결과보고서를 열람하여야 하지만, 결과보고시스템이 실제 병원 정보시스템에 통합되었기에 병원전산실에 의뢰하여 ICD-10 코드(C00-C97)로 검색하여, 환자명, 주민등록번호, 의무기록번호, 주소, 진단명, 진단코드, 최초 진단일에 대한 정보를 받았다.

중앙암등록사업단 본부로부터 1999년 1월 1일 - 1999년 12월 31일 동안 발생한 암환자에 대한 중앙암등록자료를 수집하여 상피내암 환자를 제외하고 의무기록열람을 통한 재확인 없이 악성종양

대상 및 방법

1. 연구대상

대구광역시 지역암등록사업본부(Daegu Cancer Registry, DCR)에서는 2001년 5월부터 2001년 9월까지 5개월 동안, 1999년도 대구광역시 소재 병원으로 국립암센터 중앙암등록본부(Central Cancer Registry, CCR)에 암등록을 하고 있는 6개 등록병원과 아직 등록보고를

Table 1. Cancer cases among sources (1999)

Sources	Review	Cases	Percentage(%)
CCR*	4,501 (44.7)	3,454	74.7
Pathologic reports	598 (5.9)	397	8.6
Cancer claims list in health insurance	4,463 (44.4)	436	9.4
Cancer patient list at hospital information center [†]	226 (2.3)	64	1.4
DCO [‡]	273 (2.7)	273	5.9
Total	10,061 (100.0)	4,624	100.0

* CCR : central cancer registry

[†] These are searched at hospital information center by ICD-10 code (C00-C97).

[‡] DCO : death certificates only

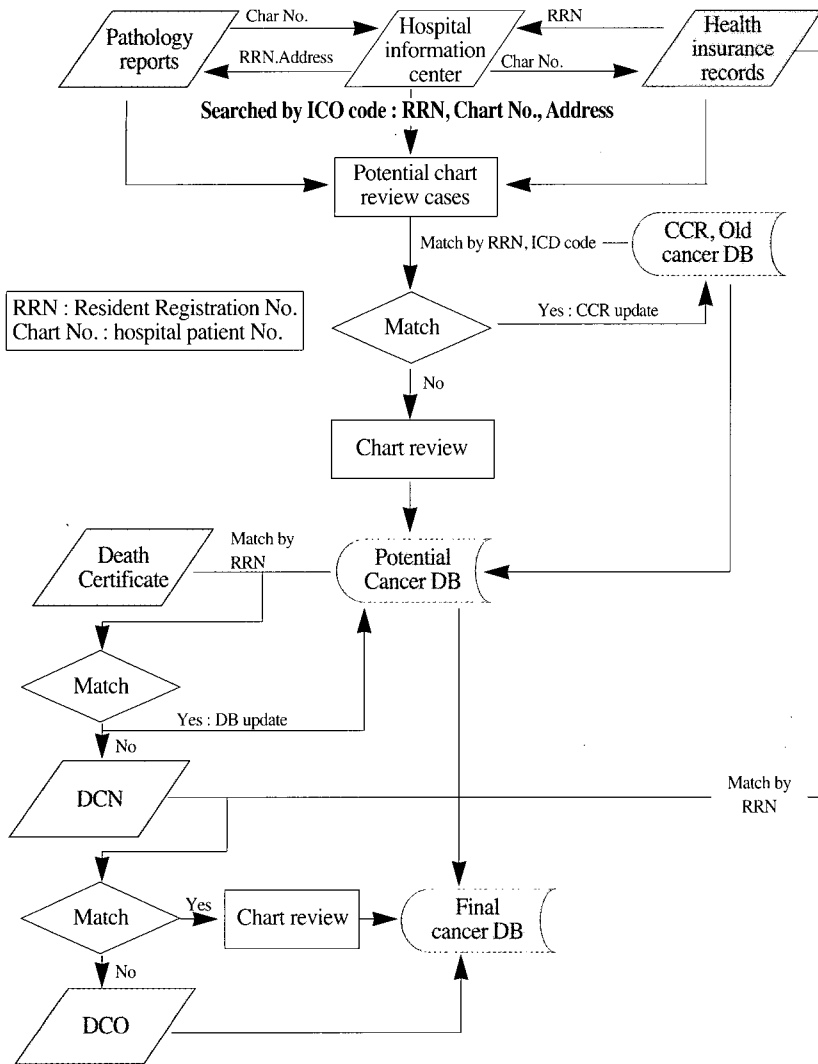


Figure 1. Diagram of cancer registration.

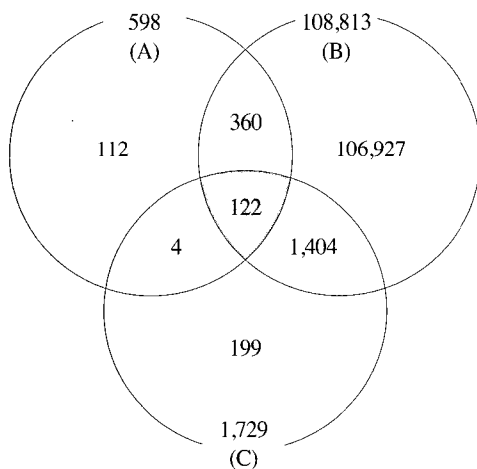


Figure 2. Matched cancer cases among sources.

만 지역암등록자료에 포함시켰다.

2) 의무기록열람

해부병리 보고서, 병원 전산실 암환자 명단에 대해 주민등록번호와 진단코드를 가지고 기존의 암환자데이터베이스 - 중앙암등록 자료(98년~99년)와 지역에서 등록된 암환자명단(96년~98년) - 와 연결하여 중복되지 않는 미등록 환자에 대해 의무기록열람을 시행하였다. 이런 후에, 의료보험공단 수진자 명단에 대해 주민등록번호를 가지고 기존의 암환자데이터베이스, 1999년 이전의 의료보험자료 그리고 앞서 열람한 명단과 중복되지 않는 환자에 대해 의무기록열람을 시행하였다.

이러한 노력에도 불구하고 암 발생자가 병원에서 진단받지 않고 사망한 경우는 누락될 수 있기에 통계청에서 사망신고서 명단을 입수하였다. 이 명단을 잠정적 암환자데이터베이스와 짝짓기(matching)를 하여 연결되지 않는 명단을 기존 암등록 자료에 등록되지 않은 Death Certificate Notification(DCN)로 확인하였다. DCN 명단을 의료보험공단 수진자 명단과 중복여부를 확인한 후, 의무기록을 열람하여 추적되지 않는 사망자를 Death Certificate Only(DCO)로 최종적인 지역암등록 데이터베이스에 저장하였다.

3) 자료 검증

최종적으로 지역암등록 자료에서 타당도를 높이기 위해, International Association of Cancer Registries (IACR)에서 제공하는 IARC Tools [6]를 이용하여 자료를 검증하였다.

4) 암등록 시스템

소규모의 연구에서는 단순한 도구(예, spreadsheets: MicroSoft Excel)로 다룰 수 있지만, 복잡한 연구는 정교한 데이터베이스를 필요로 하고, 이것은 자료 요소의 표준화, 용어정리, 입력-편집-출력을 위한 일반적인 사용환경을 제시하여야 한다 [7]. 이런 측면에서 IARC에서 제공

하는 CanReg(version 3)은 데이터베이스 개념이 없어 다중사용자환경의 대규모 자료를 다루기에는 미흡하다. 국내에서는 고려대학교 의료원 암연구소에서 병원 내 암등록 프로그램을 개발하였으나 [8], 지역암등록 사업을 고려하지 않았기에, 본 연구에서는 클라이언트-서버 기반의 입력 시스템을 새로 개발하였으며, 암정보의 접근성을 높이기 위해 인터넷기반의 암등록 시스템을 개발하였다 (Table 2).

개념적 설계(Conceptual design)에서는 업무분석(Requirement formulation and analysis)을 통해 작성된 업무기술서를 바탕으로, 개념적 설계를 위해 많이 쓰이는 개체관계모형(entity-relationship model)을 이용하여 의미모형(semantic model)을 구축하고 [9], 최종적으로 개체관계도(Entity Relationship Diagram, ERD)을 작성하였다 (Figure 3).

논리적 및 물리적 설계(Logical & physical design)에서는 ERD에 기반하여 실제 데이터베이스를 구축하였다. 암등록자료의 특성상 사망자료가 동적으로 변하는 운영계 데이터베이스가 아니라 거의 수정이 되지 않는 데이터웨어하우스 [10]의 성격을 가지기에, 환자정보레코드(person) 중 일부 항목(follow-up)은 엄밀하게 새로운 테이블로 만들지 않고, 환자신상정보에 포함하였다. 암등록레코드(cancer)에서는 다중 원발암(multiple primary cancer)을 등록하기 위해 환자 등록번호(pid)를 고유 필드(primary key)로 설정하지 않고, 암정보가 등록될 때마다 일련번호(cid)를 할당하였다. 일련번호는 8자리중, 첫 4자리는 발생년도 4자리를 표시하고 나머지 4자리는 등록된 순서대로 번호를 할당하였다. 병원레코드(hospital)에는 의료보험공단에서 제공하는 수진병원 목록(병원번호, 병원이름)을 저장하였다.

지역암등록소 내에서 쓸 수 있는 클라이언트-서버기반의 프로그램(client application)과 별도로 인터넷기반의 암등록시스템(home page)을 개발하여 암등록자료 연례보고서를 그대로 인터넷에서 내려 받을 수 있게 하거나, 보고서의

Table 2. Design of cancer registration system

Step	Detail
Platform	Server: Intel Pentium dual CPU, 256MB ram, 20GB hard disk, Linux, Oracle for linux, Apache for linux(version 1.3.22) Client: PC, MS-Window 98, Dephi(version 5)
Data modeling	Requirement formulation and analysis : description of cancer registration process Conceptual design : ERD Logical & physical design : database implementation (Record : person, cancer, hospital)
Client application	Login, on-line registration, patient search, import and export of patient list
Home page Design	Introduction of registry, annual report, web-based reporting

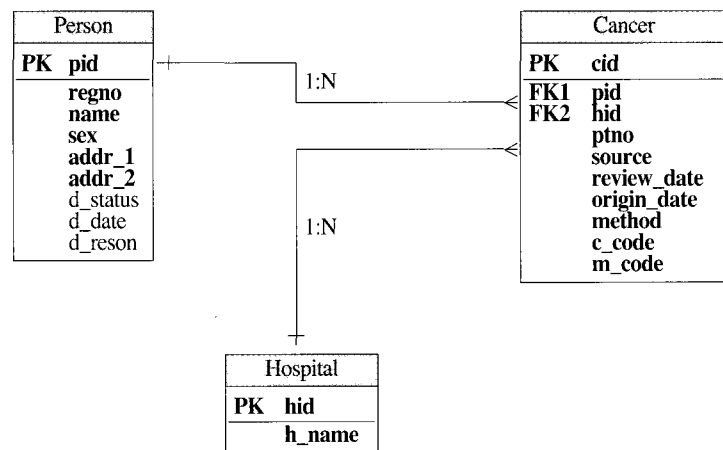


Figure 3. ERD of cancer registration.

중요 결과들을 도표 중심으로 정리해서 공개하는 보고서형태의 홈페이지를 개발하였다. 또한 지역암등록소 외부에서 표준화된 방법으로 암환자를 등록하기 위해 web-based reporting을 추가하였다.

5) 자동암등록 가능성평가

(1) 자료처리

잠정적인 1999년 발생 암환자 명단을 찾기 위해, 임상정보에 대한 접근을 허용한 2곳의 병원 전산실 암환자 명단을 기존 암환자 데이터베이스와 연결하여 중복자를 제거한다. 이 명단을 최종 지역암등록 데이터베이스와 연결하기 위해 진단코드를 IARC Tools [6]를 이용하여 ICD-10 code에서 ICD O-2 code로 변화시킨다. 이 명단과 암 데이터베이스를 주민등록번호로 서로 연결하여 전산자료 중 실제 암환자인 경우를 파악하고, 암으

로 확인된 명단에 대해서는 다시 진단코드 앞 두자리 (eg. c71.1 -> 71)에 따라 짝짓기를 하여 진단일치 여부를 파악한다.

(2) 통계적 분석방법

독립변수인 병원종류, 연령, 성(sex), 그리고 암종류가 종속변수인 진단일치에 미치는 영향을 평가하기 위해 SPSS 10.0 for Windows(SPSS Inc, Chicago, IL, USA)를 이용하여 일치도를 파악하고 범주화된 변수의 관련성은 crude odds ratio를 산출하여 비교 분석하였다. 병원종류(A와 B)는 A 병원을 reference로 하였고, 성은 여자를 reference로 하였다. 연령은 우선 10년 단위로 나뉘서 경향을 본 후에 유의한 차이가 났던 15세 이하의 소아환자를 reference로 해서, 16세 이상의 성인환자와 비교하였다. ICD O-2 code 2자리로 된 암종류 분석을 위해 15개 대분류로 나누고, 환자가 제일 많은 소

화기계암 (Gastrointestinal tract)을 reference로 하여 14개 가변수를 만들어 분석하였다. 단순 분석에서 유의한 변수들에 대해 다중 로지스틱 회귀분석을 실시하여, 변수간의 영향을 보정한 비차비 (adjusted odds ratio)를 구하였다.

결 과

1. 자료수집

자료원에 따라, 총 10,229명의 암환자를 잠정적으로 확인하고 이 명단에 대한 의무기록열람을 시행하여 발생부위, 조직형태, 주민등록번호에 따라 중복자를 제거한 후 총 4,624명의 암환자를 최종 등록하였다. 중앙암등록 자료, 사망자료 외의 새로 확인된 암환자는 897명이었다 (Table 1).

2. 클라이언트-서버 기반의 암등록 시스템

1) 사용자 인증

암등록 자료가 상당한 보안을 요하는 중요한 자료임에도 개인용 컴퓨터 사용자는 수년간 쉽게 자료에 접근할 수 있었다. 클라이언트-서버 모형에 따라 사용자는 서버 주소, 사용자명, 비밀번호를 입력하여 원격지 서버에 있는 암등록 데이터베이스에 접근한다. 이렇게 함으로써 등록된 사람만 자료를 열람하고, 또 관리자는 실시간으로 혹은 사후에 그 사람이 어떤 작업을 하는지 추적 가능하게 되었다.

2) 환자 신상정보

입력과정에서 이미 등록된 환자면 등록되었다는 안내문이 나타나고, 주민등록번호를 확인하여 입력에 오류가 있으면 오류안내문이 나타난다. 이런 다음에 서버에 접속을 시도하여 자료를 등록하게 된다. 서버와 접속 시에 오류가 생길 때는 우선 클라이언트 프로그램의 자체 데이터베이스에 자동 저장된 후 다시 접속시도를 하게 된다.

3) 암등록 정보

암등록 정보 입력은 Figure 4처럼 이루어진다.

환자 신상정보를 입력한 후에, 병원명, 발체일, 입원여부, 입원일, 환자번호를 입력한다. 다음으로, 발생일, 진단방법, 원발부위, 원발코드, 조직형태, 조직코드, 현재상태, 추적날짜를 입력한다. 이렇게 한 후, 확인 버튼을 누르면 최종적으로 암등록이 이루어지고 환자등록번호가 자동으로 할당되어진다. 입력된 내용은 아래 '목록' 창에서 다시 확인 가능하다.

3. 인터넷 기반의 암등록시스템

1) 암등록 보고서

1998년 대구광역시 암등록 사업 보고서를 참고하여, 주요한 표를 html 형식의 표로 바꾸었다.

2) 사용자 인증 및 입력

로그인을 한 상태에서 환자 신상정보

를 먼저 입력한 후, 암등록 정보를 입력한다 (Figure 5).

4. 자동암등록 타당성평가

2곳의 병원 전산실 암환자 명단(3,259명)을 기존 암환자 데이터베이스와 연결하여 중복자를 제거하여 잠정적으로 1,729명의 1999년 발생 암환자 명단을 찾았다.

이 명단(1,729명)과 지역암등록 데이터베이스의 일치도는 76.4%(1,321명)였다. A 병원의 일치도에 비해 B 병원의 일치도가 3.32배 높았고, 여자에 비해 남자가 1.64배나 일치도가 높았으며, 16세 이상에서 15세 이하에 비해 4.87배 높았다 ($p < 0.01$). 세부 질병코드별로 보면 유방암의 일치도가 91.1%로 제일 높았고, 눈·뇌 등(Eye, Brain, etc)이 35.9%로 제일 낮았다. ICD O-2 code에 대해서는

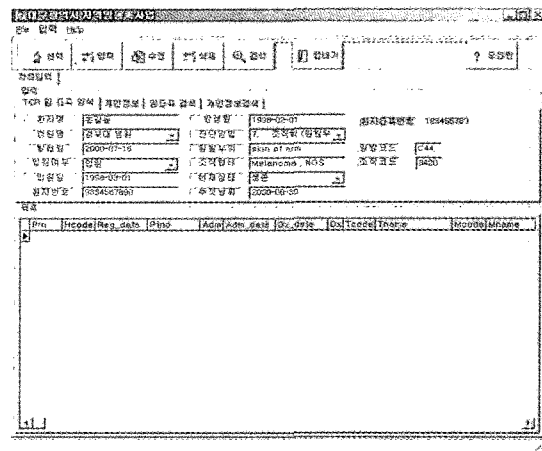


Figure 4. Input of cancer registration.

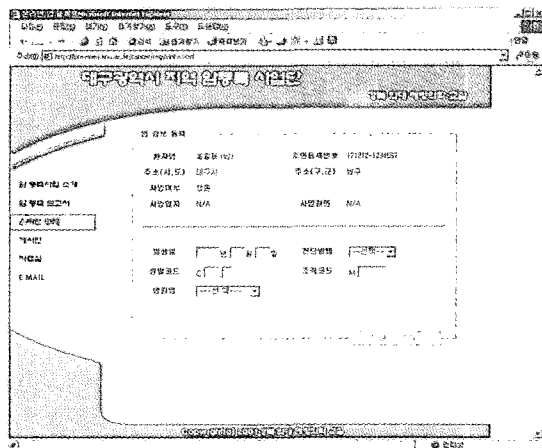


Figure 5. Input of cancer registration.

환자수가 많은 소화기계 암을 기준으로 하였을 때, 유방암의 일치도가 1.19배로 높았지만 통계적 유의성은 없었고, 눈·뇌 등(Eye, Brain, etc)이 0.07배로 제일 낮았다 ($p<0.01$) (Table 3).

병원, 성, 연령, ICD O-2 code 독립변수가 유의하여 ($p<0.01$), 질병코드 일치에 대한 병원, 성, 연령, 그리고 ICD O-2 code 영향을 보정한 비차비(adjusted odds ratio)를 파악하기 위해 다중 로지스틱 회귀분석을 시행하였다. 분석결과, A 병원의 일치도에 비해 B 병원의 일치도가 4.00배 높았고, 여자에 비해 남자가 1.62배나 일치도가 높았으며, 16세 이상에서 15세 이하에 비해 2.23배 높았다 ($p<0.05$). 세부 질병코드별로 보면 환자수가 많은 소화기계 암을 기준으로 하였을 때, 유방암의 일치도가 1.71배로 높았지만 통계적 유의성은 없었고, 눈·뇌 등(Eye, Brain, etc)이 0.05배로 제일 낮았다 ($p<0.05$) (Table 4).

고 찰

암등록 사업을 완전하게 수행하기 위해서는 실제 발생한 암 환자를 최대한 포착하여 누락되는 사람을 줄임으로써 일정 수준 이상의 완전성(completeness)을 확보하는 동시에 등록된 자료의 타당성(validity)을 확보해야 한다 [3,11].

본 연구는, 자동화된 암등록을 시도하여 지역암등록사업의 완전성을 높이고 비용을 줄이고자 시행되었다.

우선, 기본작업으로 자료원을 분석하여 동일한 자료추약 방법을 일관성 있게 적용하여 표준화한 후, 취약한 자료에서 최대한의 정보를 찾기 위해 자료원들 간의 연결을 시도하여 수동적인 열람항목을 감소시키고자 노력하였다.

본 연구에서는 주민등록번호로 연결을 시도하였는데, 주민등록번호를 제공하지 않는 해부병리자료에서는 병원전산실에 요청하거나 직접 검색하여 주민등록번호를 얻었다. 의료보험공단자료에서는 병원 의무기록번호를 제공하지 않기에 같은 방법으로 의무기록번호를 얻어야 했다. 이

Table 3. Matching between cancer patient list at hospital information center and final cancer DB among its related factors

Variables	Classification	Dismatch	Match (%)	Crude odds ratio (95% Confidence Interval)
Hospital	A	315	667 (67.9)	1.00
	B	93	654 (87.6)	3.32 (2.57-4.29)
Sex	Female	237	605 (71.9)	1.00
	Male	171	716 (80.7)	1.64 (1.31-2.05)
Age	0-15	23	16 (41.0)	1.00
	16+	385	1,305 (77.2)	4.87 (2.55-9.32)
ICD O-2 code	Gastrointestinal tract	53	456 (89.6)	1.00
	Lip, oral cavity & pharynx	13	25 (65.8)	0.22 (0.11-0.46)
	Liver & pancreas	59	176 (74.9)	0.35 (0.23-0.52)
	Respiratory & intrathoracis organs	61	217 (78.1)	0.41 (0.28-0.62)
	Bones, joints and cartilage	4	9 (69.2)	0.26 (0.08-0.88)
	Hematopoietic and reticular system	12	37 (75.5)	0.36 (0.18-0.73)
	Skin & connective, soft tissue	12	28 (70.0)	0.27 (0.13-0.56)
	Breast	15	153 (91.1)	1.19 (0.65-2.16)
	Female genital organs	86	82 (48.8)	0.11 (0.07-0.17)
	Male genital organs	14	12 (46.2)	0.10 (0.04-0.23)
	Urinary tract	13	34 (72.3)	0.30 (0.15-0.61)
	Eye, Brain, etc	25	14 (35.9)	0.07 (0.03-0.13)
	Thyroid and other endocrine glands	18	45 (71.4)	0.29 (0.16-0.54)
	Lymph nodes, etc	8	5 (38.5)	0.07 (0.02-0.23)
	Unknown primary site	15	28 (65.1)	0.22 (0.11-0.43)
	Total		408	1,321 (76.4)

Table 4. Multiple logistic regression analysis for code matching and its related factors in cancer patient list at hospital information center

Variables	Classification	Adjusted odds ratio	95% Confidence Interval
Hospital	A	1.00	
	B	4.00	2.99-5.35
Sex	Female	1.00	
	Male	1.62	1.21-2.18
Age	0-15	1.00	
	16+	2.23	1.03-4.82
ICD O-2 code	Gastrointestinal tract	1.00	
	Lip, oral cavity & pharynx	0.25	0.12-0.54
	Liver & pancreas	0.37	0.24-0.57
	Respiratory & intrathoracis organs	0.48	0.32-0.73
	Bones, joints and cartilage	0.51	0.15-1.77
	Hematopoietic and reticular system	0.54	0.26-1.14
	Skin & connective, soft tissue	0.33	0.15-0.70
	Breast	1.71	0.91-3.21
	Female genital organs	0.12	0.07-0.19
	Male genital organs	0.12	0.05-0.33
	Urinary tract	0.30	0.14-0.62
	Eye, Brain, etc	0.05	0.03-0.12
Thyroid and other endocrine glands	0.45	0.23-0.86	
Lymph nodes, etc	0.08	0.02-0.27	
Unknown primary site	0.28	0.14-0.58	

$\chi^2 = 320.40$, p -value < 0.01

런 작업을 통해 기본 의무기록 열람 자료로써 병원, 이름, 주민등록번호, 주소, 병명을 확보함으로써 의무기록 시에 기본 신상정보에 대한 확인을 간소화할 수 있었고, 자료입력 시에 다시 입력하지 않아도 되기에 오류를 줄일 수 있었다.

1997-98년 대구지역 암 발생률을 조사한 연구 [12]와 방법론을 비교하였을 때, 이 연구에서는 입원환자 이외의 환자를 찾기 위해 임상결과보고서(해부병리과, 임상병리과, 진단방사선과, 핵의학과, 치료방사선과)를 직접 열람하여 정리를 하고 환례에 식별자료가 없을 경우, 의무기록을 추적하여 주민등록번호, 주소, 발생일에 대한 정보를 수집하였지만, 본 연구에서는 전산검색이 가능한 병원은 굳이 결과보고서를 열람하지 않고 전산자료로 처리함으로써 주민등록번호, 의무기록번호, 주소, 진단명, 진단코드, 첫 진단일을 바로 확인할 수 있었다. 하지만, 본 연구에서도 진단코드를 재확인하고 발생일 - 첫 진단일이 아니라 관련증상으로 처음으로 병원을 내원한 날짜 - 을 찾기 위해 의무기록 추적을 해야 했다.

두 연구간의 완전성을 비교하였을 때, 비록 같은 연도는 아니지만 비슷한 암양상을 가진다고 가정하면, 앞선 연구의 1997-98년 2년간 환자수는 7,837명이었고 본 연구의 1999년 한해 환자수는 4,624명으로 본 연구의 환자수가 조금 많은 편이었다. 실제로, 완전성면에서 현재 지역암등록 사업에서 주요한 누락원인인 외래만 내원하는 암환자를 찾기 위해서, 하나의 방법으로 1999년 대구지역 병원의 의료보험공단 수진자 명단 108,813명을 다 열람한다고 하더라도 최종 지역암등록자료(4,624명)에서 전자와 주민등록번호로 연결되지 않는 경우가 902명(CCR 433명, 지역등록 199명)이나 되므로, 적극적으로 병원 임상결과보고서에 대한 조사가 필요한데 그 방법으로 본 연구의 검색방법이 타당하지 않을까 생각된다.

주민등록번호가 기본정보로 포함된 우리나라와는 달리, 미국에서는 자료의 기밀성 때문에 기본 정보로서 사회보장번호가 포함되지 않는 자료원에 대해 연결 프로그램으로 짝짓기함으로써 환자를 추출하는 시스템을 개발하고 있다.

미국은 기존에 구축된 환자 데이터베이스와 주정부의 사망자료, 연방정부 사회보장청(Social Security Administration)의 사망자료를 짝짓기하여 한 주의 사망자를 찾아내는 연구 [13]에서는 사회보장청 자료로부터 성(last name), 이름(first name), 출생날짜, 사회보장번호를, 주정부자료로부터는 그 변수에 middle initial, 인종, 거주지를 포함하여 얻은 후 연결 알고리즘 [14]에 따라 환자 데이터베이스와 짝짓기를 시도하였다. 의료비용에 관한 연구 [15]에서는 아이오와(Iowa)주의 SEER(Surveillance, Epidemiology, and End Results) 등록소와 입원환자의 퇴원요약지를 짝짓기함으로써 암환자의 치료, 진료비 정보를 얻었다. 또, 조기검진의 효과를 측정하기 위해 유방암 조기검진 자료와 암등록소 자료를 연결한 연구 [16]도 있다.

이렇게 암등록소에 저장된 정보를 보고서 형태로 그치는 것이 아니라, 연구를 위해 충분히 활용하기 위해서는 인터넷 기반에서 정보를 제공하여 사용자가 자신의 질의를 좀 더 쉽게 구성하게 하며, 그들이 그 즉시(just-in-time) 정보에 접근할 수 있게 해야 한다 [17].

실제 실용화한 등록소를 유형별로 정리한 천병렬과 이영성의 보고 [18]에 따르면 본 연구는 보고서의 주요 결과를 웹의 도표로 제시하였기에 가장 기본적인 형태인 '보고서형'으로 분류된다. 이후, 이 정보에 독일 소아암등록소(German Childhood Cancer Registry) [19]처럼 지역정보를 포함한다면 더 발전된 분석이 가능하고, 특히 지역주민의 생활환경 조사 정보와 연계한다면 지리정보시스템과 통합하여 전염병 감시체계처럼 암 감시체계를 구축할 수 있을 것이다. 현 시스템도 변화에 맞춰 보고서형이 아니라, 웹 기반의 등록을 보완하고, 데이터베이스에 사용자가 직접 접속하게 하여 분석 자료로서 쓸 수 있게 하거나 [20], 미국 국립 암센터 통계연구용부(Statistical

Research and Applications Branch in NCI, SRAB)의 SEER*STAT [21]처럼 기술통계분석까지 가능하게 개선할 필요가 있다.

임상정보에 대한 접근을 허용한 2개 병원의 자료로 자동 암등록 가능성을 평가하였을 때, 2개 병원의 전산실 암자료 측면에서 지역암자료와 일치도는 76.4%(1,321/1,729)이었고, 지역암자료 측면에서 2개 병원 전산실 암자료와 일치도는 28.6%(1,321/4,624)로, 이 자료만으로 자동암등록을 하기에는 부족하였다. 유럽 암등록소 네트워크(European Network of Cancer Registries, ENCR) automated cancer registration website [22]에서는 이렇게 자료원들간의 연결을 통해 자동등록을 시도하여 전체등록 환자중 수작업에 의한 등록률을 1999년에는 31%까지 줄일 수 있었다. 또 Tognazzo 등 [23]이 시행한 연구에서는 1,539명의 수작업으로 등록된 환자에서 자동등록을 시행한 결과 1,456명(94.6%)을 확인할 수 있었다. 이에 본 연구도 2곳의 병원자료뿐 아니라 다른 병원까지 전산자료 협조가 이루어지면, 지역암자료 측면의 일치도 28.6%는 더 증가될 것이다. 그리고, 병원전산실 암자료 측면의 일치도는 Tognazzo 등의 연구 [23]에서 언급하였듯이 앞으로 암 종류별로 각각의 암등록자원과 의료정보시스템에 맞는 자동등록 알고리즘을 개발해야만 증가될 것이다.

참고로, 1999년 대구지역병원의 의료보험공단 수진자 명단 108,813명을 과거 암등록자료와 연결되는 환자를 제외하고, 자료내의 중복수진자 명단을 정리한 14,868명에 대해 지역암자료와 연결하였을 때, 의료보험공단 수진자 자료 측면에서 지역 암자료와 일치도는 25.0%(3,722/14,868)이었고 지역 암자료 측면에서 일치도는 80.5%(3722/4,624 CCR 3021명, 지역등록 701명)이었다.

다변량분석에서 병원 A와 병원 B 간의 일치도 차이는 A병원에 비해 B병원의 병원전산시스템이 잘 구축되어 있고, 그 결과 환자검색 항목에서 차이가 나기 때

문이다. 소아환자에 비해 성인환자가 더 잘 일치하였으며, 암종별로 차이도 유의하였다. 부가적으로, Sex와 ICD O-2 code간의 상호작용, Age와 ICD O-2 code간의 상호작용은 유의하지 않았다. 하지만, 소아암환자가 병원전산자료에서 39명밖에 안되기에 누락된 환자가 있을 것이고, 성별차이가 나는 것으로 봐서 여성에서 호발하는 유방암, 여성생식기 암, 갑상선과 기타 내분비계 암에 대한 추가적인 검토가 필요할 것이다.

이렇게 개발된 검색 프로그램에 의해 추출된 정보는 전적으로 기저자료에 의존하기에 우리가 개발한 시스템은 이 안에 포함된 자료만큼만 가치가 있다. 즉, 자료자체가 취약하다면 사용자는 적절한 정보를 얻지 못하거나 잘못된 정보를 받게 되고 혹은 자신이 찾은 정보를 이해할 수 없게 된다 [17].

그러므로, 구축된 시스템이 효율적으로 작동하기 위해서는 우선, 각 병원 해부병리과, 의무기록실, 전산실의 적극적인 협조가 필요한데, 이미 암등록사업 경험을 통해 해부병리과와 의무기록실의 협조는 어느 정도 원활하였지만, 전산실에 대해 이런 새로운 방법에 대한 이해와 행정적, 법적인 면을 설득하는 일이 쉽지 않은 일이었다. 이런 자동화된 방법이 의무기록 열람을 줄일 수 있다고 설득하였고 또한 암관련법이 통과된다면 좀더 용이하지 않을까 생각한다. 게다가, 중앙암등록사업단에서 의료보험공단자료를 얻을 때, 의무기록번호, 주소에 대한 정보를 추가한다면 병원전산실에 추가정보를 요청할 필요가 없을 것이다. 또한, 병원마다 전자 의무기록 (Electronic Medical Record) 이 활성화된다면 좀 더 포괄적인 정보를 손쉽게 확인할 수 있게 될 것이다. 실제로, 이미 전자의무기록 시스템이 구축된 일개 대학병원에서는 컴퓨터로 검색하여 퇴원요약지를 확인하는 것만으로 현재 암등록 보고양식에서 요구하는 정보를 충분히 얻을 수 있었다. 하지만, 실제 의무기록지를 열람했을 때 퇴원요약지 정보와 실제 의무기록지의 정보가 다른 경우도 있었다. 퇴원요약지가 작성되는 시

점과 진단이 이루어지는 시점의 차이 그리고, 퇴원요약지를 무성의하게 작성함으로써 생기는 오류라고 생각된다. 그러므로, 전자의무기록이 충분한 암등록 정보를 제공하기 위해서는 병원 내에서 시스템을 구축할 시에 암등록에 관한 고려가 절실히 필요하고, 또한 입력 오류없이 정확한 정보를 제공하기 위해서는 최종사용자(의사, 간호사, 의무기록사 등)를 대상으로 요구를 적절히 반영하고 그 사용법을 교육시킬 필요가 있다 [25]. 또한, 상이한 의무기록체계에서 복잡한 암등록 사업을 수행하는 숙련된 암등록작업이 절실히 요구된다. 앞으로 암병기에 관한 항목이 기본항목으로 첨가되는 2003년부터는 이런 노력이 더욱 더 요구될 것이다.

요약 및 결론

다양한 암환자 자료원들 간의 연결을 통해 암등록사업의 완성성을 높이고, 그 자료를 데이터베이스로 저장하여 인터넷을 통해 암환자를 등록하고 등록결과를 공유함으로써 기초암연구 자료로서 지역 암등록 자료를 유용하게 이용 가능하게 하는 전산시스템을 개발하며, 자동화된 암등록의 기초작업으로 현재의 병원전산자료에서 검색한 암환자명단이 암등록자료로 가능한 지를 알아 보고자, 2001년 5월부터 2001년 9월까지 5개월 동안, 중앙암등록자료, 해부병리결과 보고서, 의료보험공단 수진자 명단, 병원전산실 암환자명단, 사망진단서로부터 1999년도에 암으로 보고한 10,061명의 환자를 대상으로 조사를 하여 4,624명의 대구지역 암환자를 확인하였다. 그 중 실제 지역에서 의무기록 열람을 통해 확인한 암환자는 897명이었다. Cross record linkage를 통해 암환자를 좀더 효율적으로 찾고 입력오류를 줄일 수 있었다. 암등록자료를 보고서 형태로 제출하는 것에 그치는 것이 아니라 기초암연구에서 활용하기 위해서 보고서형태의 인터넷 홈페이지를 개설하며 원활한 등록을 위해 클라이언트-서버 모델 그리고 웹기반의 암환자등록 시스템을 구현하였다.

임상정보에 대한 접근을 허용한 2개 병원의 자료로 자동암등록 가능성을 평가하였을 때, 일치도는 병원, 연령, 암종별로 유의한 차이가 있었다. 2개 병원 전산실 암자료 측면에서 지역암자료와 일치도는 76.4%이었고, 지역암자료 측면에서 2개 병원 전산실 암자료와 일치도는 28.6%로 이 자료만으로 자동암등록을 하기에는 부족하였는데 2곳의 병원자료뿐 아니라 다른 병원까지 전산자료 협조가 이루어지면 지역 암자료 측면의 일치도 28.6%는 더 증가될 것이다. 향후 좀 더 적극적인 병원 전산실의 협조가 필요하며, 암 종류별로 각각의 암등록자원과 의료정보시스템에 맞는 자동등록 알고리즘을 개발해야만 할 것이다.

참고문헌

1. 통계청. 주요사망원인 사망률. 2002. Available from: URL: <http://www.nso.go.kr/>
2. WHO Statistics. Numbers of deaths and death rates. 2002. Available from: URL: http://www3.who.int/whosis/whsa/whsa_table1.cfm?path=whosis,whsa,whsa_table1&language=english
3. Shin MH, Ahn YO. Evaluation of the completeness and validity of the registration in the implementation study of Seoul Cancer Registry. *Korean J Prev Med* 1994; 27(4): 735-745 (Korean)
4. 정규원. 1999년 부산지역 중앙암등록 충실도조사. 2001년 암등록·통계 심포지엄 자료집. 국립암센터; 2001, (125-138쪽)
5. 천병렬, 감신, 이경수, 이주영. 대구광역시 지역 암등록사업 보고서(1999년 암환자). 경북대학교 의과대학 예방의학교실; 2001
6. IACR. IARC tools. 2001. Available from: URL: <http://www-dep.iarc.fr/resour/software/iarctools.htm>
7. Prakash M, Nadkarni, Cynthia Brandt, Sandra Frawley, Frederick G. Sayward, Robin Einbinder, et al. Managing attribute-value clinical trials data using the ACT/DB client-server database system. *J Am Med Inform Assoc* 1998; 5(2): 139-151
8. Lee KN. Cancer registry program at the cancer institute in Korea university medical center. *J Korean Soc Med Inform* 1999; 5(1): 125-130 (Korean)
9. 문송천. dataware 설계론. 아이포스트; 1999, (21-116쪽)

10. Alex B. Building data mining applications for CRM. McGraw-hill; 1999. p. 16-40
11. Parkin DM, Muir CS. Comparability and quality of data. Cancer incidence in five continents Vol. 6, IARC : Lyon; 1992. p. 45-55
12. Lee CW, Lee MY, Lim HS, Sohn SS, Jeon JK. Cancer incidence in daegu in 1997-98: the first results of the daegu cancer registry. *J Korean Cancer Assoc* 2001; 33(2): 136-148 (Korean)
13. Thomas B. Newman, Andrew N. Brown. Use of commercial record linkage software and vital statistics to identify patient deaths. *J Am Med Inform Assoc* 1997; 4(3): 233-237
14. Jaro M. Probabilistic linkage of large public health data files. *Statistics in Medicine* 1995; 14 :491-498
15. Brooks JM, Chrischilles E, Scott S, Ritho J, Chen-Hardee S. Information gained from linking SEER cancer registry data to state-level hospital discharge abstracts. *Med Care* 2000; 38(11): 1131-1140
16. Schouten LJ, de Rijke JM, Schlangen JT, Verbeek AI. Evaluation of the effect of breast cancer screening by record linkage with the cancer registry. *J Med Screen* 1998; 5(1): 37-41
17. Alexa T. McCray, Nicholas C. IDE. Design and implementation of a national clinical trials registry. *J Am Med Inform Assoc* 2000; 7(3): 313-323
18. 천병렬, 이영성. 암등록통계의 실용화 프로젝트. 2000년 암등록·통계 심포지엄 자료집. 국립암센터; 2000, (89-134쪽)
19. German Childhood Cancer Registry. Geographical analyses. 2002. Available from: URL: http://info.imsd.uni-mainz.de/K_Krebsregister/english/
20. IARC. Globocan 2000. 2002. Available from: URL:<http://www-dep.iarc.fr/globocan/globocan.html>
21. Statistical Research and Applications Branch in NCI (SRAB). SEER*STAT. 2002. Available from: URL:<http://srab.cancer.gov/>
22. ENCR Automated Cancer Registration website. Improving the Algorithm for Automated Registration. 2002. Available from: URL: <http://159.92.82.173/encr/ENCR%20Workshop/Chris/Venice%20Chris.ppt>
23. Tognazzo S, Andolfo A, Fiore AR, Greco A, Guzzinati S, et al. Quality control of cancer cases automatically defined by the automated registration system of the venetian tumour registry. IACR 2002 24th Annual Meeting of the International Association of Cancer Registries Abstract Book 2002; 79
24. Hwang JI, Park HA. Nurses' attitude toward computerization and their need assessment in pre and post hospital information systems. *J Korean Soc Med Inform* 2001; 7(1): 57-66 (Korean)