

## 자동코드화 방식을 이용한 과거검사 10만 증례의 전산화 경험

서울대학교 의과대학 병리학교실

김우호 · 최기영 · 서정욱 · 김용일

### 머 리 말

외과병리 자료를 이용한 연구에서 연속 증례를 수집하는 경우 연구의 초기 단계부터 객관적인 선택 기준에 의하여 빠짐없이 모으는 것이 중요하다. 장부나 보고서 사본에서 해당 증례를 직접 찾거나 또는 코드에 의해 분류해 놓은 file card를 이용하는 증례의 방법으로는 이러한 증례수집의 원칙을 만족시킬 수 없을뿐 아니라 많은 시간과 노력을 요하는 문제점이 있었다. 서울대학교병원 병리과에서는 1978년도부터 매증례에 ICD(international classification of disease) 코드를 부여하여 1년마다 같은 번호끼리 수동식으로 정리한 카드를 제작 보관하였다. 그러나 ICD 질병분류는 장기(topography)를 위주로 한 단일축 분류(single axis code)로서 종양의 조직학적 분류, 특정한 형태의 염증, 희귀한 질병 등이 구체적으로 코드화되지 못하는 단점이 있다. 예를들면 폐의 악성종양이 모두 같은 코드로 저장되기 때문에 연구용으로 증례를 찾을 때 제대로 찾기 힘들어 실제 사용은 매우 제한되었다. 그후 1981년도부터 진단의 코드를 SNOP(systematized nomenclature of pathology) 체제에 의해 같은 방법으로 카드화 하였는데, SNOP은 다축분류법(multi-axial system)으로서 장기와 형태학적 진단이 별도로 코드화되므로 검색용으로 사용할 수 있는 범위는 크게 넓어졌다. 그러나 코드화가 번거로움에도 불구하고 수작업에 의한 코드화의 한계로 인해 연구용으로 증례를 모으는 일은 역시 제한될 수밖에 없었다\*. 즉 코드집에서 쉽게 찾을 수 없는 진단명이나 희귀한 질병은 제멋대로 코드화되거나 광범위한 의미의 코드로 기록되어 있어서 많은 노력을 들여 코드화 했

음에도 불구하고 실제 이용가치는 그리 크지 않았다. 예를 들어 담낭의 malakoplakia를 찾으려면 담낭의 만성염증 증례 모두를 확인해야만 하였고 소결절성 간경화증은 단순히 간경화증으로 분류되었으며, 급성간염은 급성염증과 급성바이러스성 간염의 두가지로 코드화 되는 등의 문제점 등이 발견되었다.

1978년부터 약 10년간 환자 이름의 가나다 순으로 정리한 카드를 별도로 만들어 정리를 하였으나, 증례가 십만 개를 넘으면서부터 카드의 양이 엄청나고, 매일 매일 새로 추가되는 카드를 제자리에 꽂는데 만도 많은 시간이 소요되었다. 더구나 흔한 이름의 환자는 수백명이 등록되어 있어 이 카드로는 하루 100여명에 달하는 병리학적 검사 의뢰 환자의 과거검사를 모두 찾아내기가 대단히 어려워졌다. 따라서, 외과병리 자료의 체계적 보관과 연구용 증례의 효율적인 수집을 위한 전산화가 절실히 필요하였다.

### 본론 및 고찰

1988년 후반기에 자동코드화 방식의 전산화를 시도하여 1989년 중반기부터 이를 본격적으로 사용하였으며<sup>2)</sup>, 1989년 증례부터는 연구용 증례의 수집이 매우 쉬워지고 환자의 과거 검사가 자동적으로 검색되는 등의 개선이 있었고, 이에 따라 과거의 기록을 소급하여 전산화할 계획을 세우게 되었다. Hardware는 IBM PC386 호환기종(80386 processor)인 삼보회사 제품(Trigem)에 7 bit 조합형 한글카드, 80 MB hard disk를 갖추었고 EPSON LQ-2550 프린터를 연결하였다. Software는 Disk Operating System (IBM) 3.30와 dBASE III plus(Ashton tate)를 기본으로 제작한 외과병리 전산화 프로그램(SURGCOM)을 사용하였다. 서울대학교 병원 병리과의 전산화 프로그램(SURGCOM)은 46개의

\*본 논문의 요지는 1990년도 대한병리학회 춘계 학술대회에서 전시 되었음.

업무단위로 되어 있고, 철저히 menu 방식으로 되어 있으며 자료를 두개의 파일에 저장토록 되어 있다. 그중 하나는 진단이 나가기 전까지 자료를 일시적으로 저장하는 임시파일로서 보고서에 타자되는 모든 항목이 저장되어 있어 1증례당 6,000 byte를 차지한다. 또 하나의 파일에는 중요한 정보만을 영구적으로 저장하는데 여기에는 병리번호, 병록번호, 이름, 성별, 연령, 진단코드가 수록되어 있다. 환자의 보고서를 타자하면 증례는 자동적으로 영구파일에 저장토록 되어 있으며, 별도로 영구 파일에 자료를 입력할 수도 있었다. 과거 검사기록의 입력은 제본된 보고서 사본을 이용하였는데, 모든 진단이 그대로 자동코드화 되기 어려우므로 진단중 코드화 되어야 할 부분에 밑줄을 표시하였다. 입력은 통상 외과병리 업무에 지장을 주지 않도록 전문 용역 업체에 의뢰하였는데 1증례당 입력 비용은 20원 정도이었다(Table 1). 컴퓨터에 입력된 자료는 디스켓에 복사하여 전달받았으며 이를 하드 디스크에 옮겨 자동코드화를 시행하였다. 하드 디스크는 DOS 3.30을 사용하여 30MB를 단위로 나누었는데 10만 증례가 차지하는 용량은 18MB이었으며 DOS 4.0을 이용하여 80MB를 한꺼번에 사용하면 40만 증례를 저장할 수 있으리라 기대된다. 지난 1년 동안 사용중 software나 hardware의 failure는 없었으며, 증례의 추가로 인해 작동시간의 길어진 몇가지 업무는 프로그램을 보완하였다.

자동코드화용 진단명 사전은 SNOMED (Systematize nomenclature of medicine) microglossary for surgical pathology를<sup>3)</sup> 모두 입력한 후 본원에서 지난 1년간 사용하였던 진단명 약 1,000 단어를 추가하여 약 3,000개의 진단명을 포함하고 있다. 자동코드화는 컴퓨터에 수록된 진단명 사전에 있는 단어만 코드화되므로

철자가 틀리거나, 컴퓨터에 수록되어 있지 않은 진단명을 사용하는 경우에는 코드화되지 못하고 진단명이 그대로 남아 있었다. 일단 자동코드화 작업을 한 후 코드화되지 않은 증례들의 진단명을 수정하여 다시 코드화를 시행하였는데 SNOMED 원본에는<sup>4)</sup> 46,000개의 진단명이 수록되어 있으므로 이들을 추가함으로써 99% 이상의 증례를 코드화할 수 있었다. 동의어인 경우에는 진단명 사전에 진단명을 같은 코드로 추가하였으며 SNOMED 원본에도 수록되어 있지 않은 진단명은 상위 개념의 진단명을 추가하여 코드화시켰다(Table 2). 이러한 작업 후에도 약 1%의 증례는 코드가 완전하지 않아 다시 입력하였던 바, 대부분 밑줄을 잘못 표시한 증례이었다.

지난 1989년 7월부터 이 작업을 시작하여 현재 7년치 (100,000 증례)를 전산화 하였는데, 1990년도에도 계속 하면 20년치를 전산화 할 수 있으리라 기대된다. 현재 정리된 100,000 증례는 장기별, 진단별로 숫자를 헤아려 목록집을 만들었으며 (Fig. 1) 필요한 경우 해당 증례의 목록을 인쇄하거나 diskette에 복사할 수 있도록 되어 있다.

자동코드화된 자료를 이용한 연구는 연구자의 병리학적 재검색을 전체로 하기 때문에 악성종양의 분화도나 세분류는 자동코딩에서 제외하였으며, 진단이 두개 이상인 경우에는 각 진단에 증례 숫자가 더해지도록 하였다. 예를 들어 악성 림프종의 세분류나 선암종의 분화도 등은 자동코딩에서 제외하고 진단명만을 입력하였으며, 대결절성 간경변증의 증례 숫자에는 간세포암과 동반된 증례가 상당수 포함되어 있었다.

1990년도 3월부터 이 자료를 이용하여 연구용 증례의 검색에 이용하고 있으며 몇가지 증례에 대해 비교한 결

**Table 1.** 외과병리 검사결과 전산화시 소요되는 시간 및 비용 (10,000 증례당)

작업	작업자	시간	비용
진단명에 밑줄긋기	전공의	40시간	20만원
컴퓨터에 입력하기	타자수 (용역업체)	(1주일)	
자동코드화	컴퓨터	5시간(night)	
잘못된 코드 수정하기	전공의	4시간	
컴퓨터 자료 수정하기	타자수	(2일)	
자료 인쇄 (진단별 증례 수)	컴퓨터	2시간	

**Table 2.** 자동 코드화시 코드 수정방법의 예

분 류	예		전체 증례에 대한 비율
	수정전	수정후	
1. 철자가 틀림	stmach	stomach	15%
2. 동의어	femoral head	femur head 와 같은 코드로 dictionary에 추가함	2%
3. SNOMED 원본에 있는 진단명	ovotestis	M23140 를 dictionary에 추가함	2%
4. SNOMED 원본에도 없는 진단명	von-Gierke's disease	glycogen storage disease Darier's disease	3%

<< T56 = >> liver		<< M95903 = >> malignant lymphoma	
M00100 = normal tissue	44	T01000 = skin	20
M09000 = insufficient material for diagnosis	18	T04000 = breast	1
M09450 = no evidence of malignancy	1	T07000 = spleen	12
M14400 = laceration	1	T08000 = lymph node	123
M20000 = congenital anomaly	4	T11710 = femur	2
M30000 = calculus	64	T12000 = joint	1
M30010 = cholelithiasis	8	T1X000 = soft tissue	9
M33060 = bile stasis	15	T21000 = nose	18
M33400 = cyst	9	T22000 = paranasal sinus	2
M33401 = simple cyst	3	T23000 = nasopharynx	7
M33700 = parasitic cyst	1	T24100 = larynx	2
M34000 = obstruction	9	T29000 = pleura	3
M34060 = bile duct obstruction	1	T31000 = pericardium	1
M35100 = thrombus	3	T48000 = vein	1
M35300 = embolus	1	T51120 = hard palate	7
M36142 = chronic passive congestion	9	T52000 = lip	1
M37100 = hematoma	2	T53000 = tongue	4
M40000 = non-specific inflammation	43	T54910 = gum	2
M40059 = reactive hepatitis	47	T55100 = parotid gland	3
M41000 = acute inflammation	8	T56000 = liver	4
M41400 = acute suppurative	2	T60200 = oropharynx	5
M41740 = abscess	24	T61100 = tonsil	30
M42000 = subacute inflammation	7	T63000 = stomach	77
M42100 = acute and chronic inflammation	2	T64300 = duodenum	7
M42101 = chronic active inflammation	37	T65200 = ileum	14
M42700 = subacute necrotizing inflammation	1	T66000 = appendix	1
M43000 = chronic non-specific inflammation	38	T67000 = colon	22
M43700 = chronic necrotizing inflammation	1	T68000 = rectum	6
M43740 = organized abscess	2	T71000 = kidney	1
M44000 = chronic granulomatous inflammation	7	T78000 = testis	6
M44140 = foreign body reaction	1	T96000 = thyroid	5
M45000 = fibrosing inflammation	1	TX2200 = frontal lobe	3

Fig. 1. Example of count menu. Left column lists up cases in the liver and right column displays cases of malignant lymphoma.

과 보관된 file card나 장부를 뒤져서 찾아낸 것보다 훨씬 쉽게 그리고 더 많이 찾을 수 있었으며 종래 방식으로 선택된 자료의 중복과 누락을 확인할 수 있기도 하였다. 또한 이렇게 저장된 환자 기록은 같은 환자의 생검이 다시 의뢰될 때 육안기록의 타자시 함께 인쇄되어 병리의에게 전달되므로 지난 7년간의 환자의 과거 기록을 자동적으로 참조할 수 있게 되었다. 따라서 여러 장기의 복합적인 질병을 시사할 두고 겪어온 환자의 경우, 불충분한 임상정보로 인해 각 장기의 질환 개개를 독립적으로만 진단하는 것에 그치지 않고, 환자의 전반적인 질병상태를 이해할 수 있게 되었다.

SNOMED를 이용한 자동코드화 방식은 외과병리 자료 정리에 매우 편리하고 유용한 방식으로 생각된다. 현재 이 프로그램은 서울대학교 병원 이외 7개 이상의 기관에서 효과적으로 사용하고 있으며 이 프로그램을 변형시켜 소화기내시경<sup>5)</sup>, 골수검사<sup>6)</sup>, 방사선과<sup>7)</sup> 등에서 적용한 프로그램들도 소개되어 있다.

## 참 고 문 헌

- 1) Ganter GE: Chapter 23. Coding aid. In *systematized nomenclature of medicine coding manual*. Ganter GE (ed) Illinois, College of American Pathologists. 1979, p 49
- 2) 김우호, 서정옥, 김용일: 개인용 컴퓨터를 이용한 자동코딩 방식의 진료업무 전산화. *대한병리학회지* 23: 410-415, 1989
- 3) Rothwell DJ: *Systematized nomenclature of medicine. Microglossary for surgical pathology*. Illinois, College of American Pathologists, 1980
- 4) Cote RA: *Systematized nomenclature of medicine. Vol II (Alphabetic index)*. Illinois, College of American Pathologists, 1982
- 5) 김우호, 정현채, 송인성: 자동코드화 방식을 적용한 소화기 내시경 검사보고서 작성 및 자료정리 업무 전산화. *대한소화기내시경학회잡지* 10(1):95-103, 1990

- 6) 윤영호, 김우호, 최영식, 정화순 : 개인용 컴퓨터를 이용한 자동코딩방식의 골수검사 및 말초혈액도말상 보 고서의 전산화 대한임상병리학회지 10:335-342, 1990
- 7) 김우호, 이병희, 이경상, 한준구, 최병인, 한만청 : 미국방사선학회코드(ACR code)의 자동코드화 방식에 대한 검토, 대한방사선학회지 26(2):432-438, 1990

= Abstract =

### Computerization of 100,000 cases of Surgical Pathology Data at SNUH by Automatic Coding System using Personal Computer

Woo Ho Kim, M.D., Ghee Young Choe, M.D.  
Jeong Wook Seo, M.D. and Yong Il Kim, M.D.

*Department of Pathology, College of Medicine,  
Seoul National University*

A computer program using automatic coding of the diagnosis has been used for report printing as well as data storage and retrieval system at the Department of

Pathology, Seoul National University Hospital. Previous surgical pathology files were also computerized by the automatic coding system using personal computer, and 100,000 cases of surgical pathology data during 7 years were computerized at present.

The computerized surgical files were counted and listed according to topograph and morphologic diagnosis. It is available to print out the list of a specific diagnosis or to copy the records to a floppy diskette.

Collection of cases in surgical pathology files using computerized automatic coding system becomes much convenient and accurate than using stored file cards or log books. In addition, previous biopsy records of the patient are automatically searched during the routine work so that understanding of a patient as a whole is possible through the informations about previous pathologic diagnosis. We confirm that automatic coding method is the most practical and economic method for computerization of the surgical pathology records.

**Key Words:** Computerization, Surgical pathology, Automatic coding system, Personal computer