AUTOMATIC TNM STAGING OF CANCER
Using the VA-DHCP ONCOLOGY-V2 TUMOR REGISTRY

Susan R. Richie
Department of Veterans Affairs
Washington Information Systems Center

ABSTRACT

Cancers can be subdivided into uniform groups, or TNM stages, by evaluating the size or extent of the primary tumor (T), the presence or absence of regional lymph node involvement (N), the presence or absence of sites of distant metastases (M), and combining them together into a stage using a simple classification scheme. Standard classification and staging of cancer allows the physician to determine treatment more appropriately, to evaluate case management results more accurately and make credible statistical comparisons to other institutions. The practice of dividing cancer cases into groups according to "stages" arose from the fact that survival rates were higher for cases where the disease was localized compared to cases where the disease had extended beyond the organ or site of origin, referred to as the "primary site". The VA DHCP (Decentralized Hospital Computer Program) Oncology Version 2 automates this staging process for all sites for which a staging scheme has been devised, providing consistency in data collection, interpretation, comparisons, and statistical analysis.

INTRODUCTION

The DHCP-Automated Tumor Registry for Oncology Package, Version 2 presents significant enhancements over Version 1.1. Programmed in standard MUMPS and the VA File Manager, it is powerful, fast and comprehensive. One programmer worked with thirty test sites, producing a very user-oriented product, filled with the input of experienced VA registrars. This is truly a registry package created by registrars for registrars: software developed in concert with the user.

Registries developed as an outgrowth of treatment research and the need to maintain long term follow-up information for use in statistical analysis. Historically, the task of searching for qualified cases for inclusion, the tracking of these cases until properly abstracted, following up annually and then extracting the data out for statistical analysis expended innumerable man-hours and covered thousands of pages of paper. A large staff was needed to sift through these mountains of paperwork. Valuable and often unavailable office space was needed to house not only the staff but the file cabinets required to keep the records.

With President Nixon's declaration of war on cancer in the 1970's came the need for data collection standardization. The American College of Surgeons (ACOS), and the American Joint Commission on Cancer (AJCC) became the regulating bodies. During this same time frame, the computer became the method of choice for storing data. The first attempts at developing computerized registries were nothing more than modified commercial application database programs. During this time frame, the ACOS developed and started marketing its own software, CanSurNet, to meet its own data gathering standards published in the ACOS Data Acquisition Manual. On its heels came other ventures, mostly stand alone packages running on personal computers and basically meeting the standards of the ACOS. The drudgery of collecting and manually entering the information continued, although maintenance and analysis became much easier.

Oncology Version 2 was designed with a primary goal in mind: to eliminate all the drudgery of maintaining a registry including double entry of data; a good part of this has been accomplished, but much more remains to be done. Created for use within the VA system, the finished product has the advantage over commercial packages: working in a DHCP environment, the package has extensive links to other DHCP products and access to data already existing in the integrated computer system, allowing the incorporation of several unique features.

Automatic Casefinding is the first and most impressive of these unique features stemming from an integrated system. By accessing the Casefinding and Suspense module, the user may specify a time frame for the computer to search the Laboratory, Anatomical Pathology, Radiology, and MAS Patient Treatment file (PTF) for cases which meet the established criteria for registry inclusion. The cases found are automatically entered into a suspense file, along with relevant data, where they are held until abstracted into the registry.

Abstracting is the second most important aspect of registry work. Strict adherence to coding standards and the requirements of the regulating bodies is necessary for accreditation. Abstracting becomes effortless as the package guides the registrar through the abstract prompting for detailed information that meets the standards and requirements every step of the way. Help screens are available for every data element simply by typing a '?'. Demographic data already existing within DHCP, automatically becomes part of the Oncology database or is accessible to it through computed fields, thus eliminating labor intensive double entry of data.

To assist the registrar in astracting, the package incorporates following references: the ACOS Data Acquisition Manual- 1990, including the SEER (Surveillance, Epidemiology and End Results) Extent of Disease 1988 coding manual, AJCC Manuals for Staging of Cancer,(3rd and 4th editions), the International Classification of Diseases for Oncology, 1st and 2nd Editions, and the International Classification of Diseases, Clinical Modification, 9th Revision.

The package contains the following modules:

- o Casefinding and Suspense
- o Abstracting and Printing
- o Follow-up
- o Registry Lists
- o Annual Reports
- o Statistical Reports
- o Utility Options

Follow-up is unique in this program as both the cancer patient, and the cancer case, or primary, can be followed. For each patient follow-up there is an associated Tumor Status follow-up. This is important in the situation of a patient having multiple primaries, or more than one site of origin: if the patient has no evidence of cancer, then the tumor status is set to no evidence for each primary. However, if the patient shows evidence of cancer, or the cancer status is unknown, then the user is prompted for the tumor status for each primary, and the current patient cancer status, as well as the current tumor status for each primary automatically is updated. This method saves the Tumor Registrar much time in carrying out the follow-up, not having to repeat data entry for each primary, and allows for tracking of the whole patient, as well as each cancer case, making it more patient oriented - unusual for Tumor Registry software.

Reporting to the central registries is available using the File Manger options to either print the data to paper or to diskette. Reporting the ACOS for their yearly Call for Data is accomplished through the IRM site manager option that extracts the required data elements, formats them according to the ACOS rules, then saves the data to disk for mailing.

The statistical module produces Two-way cross tabulations, and using computed fields, many combinations can be quickly be produced showing interesting relationships, without exporting data into a spreadsheet package. Attachment #3 illustrates the data for a ten year period between selected sites and stage grouping. The data can be easily transported to graphics software. Actuarial lifetables and crude survival curves are also available.

In summation: the package is easy to use, comprehensive, powerful, full featured, cost effective, and compliant to standards. In these tight budget times, DHCP Automated Tumor Registry for Oncology Version 2 is an economical choice.

AUTOMATIC TNM STAGING

The American Joint Commission on Cancer (AJCC) first published a "Manual For The Staging Of Cancer" in 1977. Four hundred participants over 25 years, including retrospective studies of literature and exhaustive meetings with the International Union Against Cancer (AJCC), led to the evolution of the staging system in use today. To further assist the physician and the registrar Oncology V2 introduces automatic staging for all primary sites. The idea of automatic staging was introduced by Dr. Margaret Fletcher and Chris McManus in the ENT Tumor Registry innovative software described in 1987 (1). Also written in MUMPS and the VA File Manager, was limited to staging the head and neck anatomic sites.

The staging of cancer is not a fixed science: the consistency in "Manual for the Staging of Cancer" was to compile all available data on the staging of neoplastic diseases in different organs and systems. The importance of data uniformity cannot be overstressed. Quality and consistency in data are requirements for effective treatment planning, and meaningful survival analysis. By providing staging forms for staging each body system, with checklists to assist the clinician with determining the extent of disease, uniformity was achieved (see attachment #1). In 1992 the AJCC published their 4th edition, modifying many of the T,N,M descriptions, refining and increasing the gradient definitions, changing the algorithms for a number of sites, and adding staging schema for ones previously undefined.

The system is a shorthand notation for the describing the clinical extent of a particular malignant tumor. Clinical classification, designated CTNM of TNM is based on all evidence acquired before any definitive treatment from physical examination, imaging,endoscopy, biopsy, surgical exploration and other relevant medical procedures. Pathological classification, designated as PTNM includes all of the above, supplemented or modified by additional evidence acquired from surgery, and from pathological examination. Once the T,N,M and/or the Pt,Pn,Pm is assigned and the stage determined and recorded in the medical record, it must remain unchanged. With a computerized system this has obvious implications which do need to be addressed in the future.

The DHCP Oncology Tumor Registry package currently allows for either using the clinical or pathological schema designated by the staging basis field. It would be desirable in the future to have the ability to have both, and this probably will become a ACOS requirement. All the T,N,M descriptions for all defined sites, both for the 3rd and 4th staging editions, and the toggle for switching between the two resides in the Site Parameters file. Each staging table or algorithm associating a T,N,M to with a particular stage, reduces to a Boolean expression written in MUMPS.

To accomplish the total automatic staging system, a series of 10 programs were written, and two data files populated with the T,N,M coding data with descriptions used in the diagnosis of cancer. Originally the design was to put the staging tables in a file, using a cross-reference as a table; however due to time constraints, and the future changing of a the algorithms, mumps code was used. MUMPS has such great flexibility that it was easier to add the fourth edition than it might have been in other languages. The Automatic Staging process in the Version 2 package helps to eliminate these errors introduced when combining the T,N,M variables into a stage for a particular anatomic location. Once a primary site is chosen, only site-specific information is displayed for entry into the system. The field that serves as a switch in the site parameters file, defines when the registry switched from the 3rd to 4th edition, 1992 or 1993. The program looks at the accession year of the case, the decision year, and displays the appropriate T,N,M codes

for that primary site, either from the 3rd or the 4th edition depending upon the accession year, and the year the registry switched editions.

Prostate TNM coding illustrates the differences between the 3rd and 4th staging editions as Prostate had the most significant revision.

AJCC TNM STAGING: PROSTATE
**** Indicates 4th Edition additons

TUMOR:

TX		Primary tumor cannot be assessed
T0		No evidence of primary tumor
T1		Clinical: inapparent not palpable/visible by imaging.
****	T1a	Incidental finding in 5% or less of tissue resected.
****	T1b	Incidental finding in more than 5% of tissue resected.
****	T1c	Identified by needle biopsy (e.g., due to elevated PSA)
T2		Tumor confined within the prostate
****	T2a	Tumor involves half of a lobe or less both lobes
****	T2c	Tumor involves both lobes
T3		Tumor extends through the prostatic capsule
****	T3a	Unilateral extracapsular extension
****	T3b	Bilateral extracapsular extension
****	T3c	Tumor invades the seminal vesicle(s)
T4		Fixed/or invades adjacent structures other than vesicle(s)
****	T4a	Invades: bladder neck, external sphincter, or rectum
****	T4b	Invades levator muscles and/or fixed to pelvic wall

LYMPH NODES:

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in single lymph node; 2cm or less
N2	Metastasis in single lymph node; 2cm to 5 cm
N3	Metastasis in single lymph node; >5cm

METASTASIS:

MX		Distant metastasis can't be assessed
M0		No distant metastasis
M1		Distant metastasis
*****	M1a	Nonregional lymph node(s)
*****	M1b	Bone(s)
*****	M1c	Other site(s)

GRADE: Differentiation

GX	Can't be assessed
G1	Well differentiated
G2	Moderately well
G3-4	Poorly/undiff.

All four variables will go into staging the cancer as shown in the table.

STAGE GROUPING of PROSTAGE - 4th Edition

Stage 0	T1a	N0	M0	G1
Stage 1	T1a	N0	M0	G2,3-4
	T1b	N0	M0	Any G
	T1c	N0	M0	Any G
	T1	N0	M0	Any G
Stage II	T2	N0	M0	Any G
Stage III	T3	N0	M0	Any G
Stage IV	T4	N0	M0	Any G
	Any T	N1	M0	Any G
	Any T	N2	M0	Any G
	Any T	N3	M0	Any G
	Any T	Any N	M1	Any G

The table translates to the following MUMPS expression:

```
;PROSTATE-C61.9: 4th Edition
S SG=$S(M!N!(+T=4):4,+T=3:3,+T=2:2,T'["A"!(G>1):1,G:0,1:"E") Q
;
;PROSTATE: 3rd edition expression
S SG=$S(M!N!(T=4):4,T=3:3,T["B":2,G>1:1,+T:0,1:"E") Q
```

In the VA, clearly 75% of the tumors diagnosed are prostate related, and staging is more complicated. The AJCC coding scheme requires that the grade of the tumor be accounted for in the determination of stage. The staging algorithm will look first at the primary site, the grade, the T, the N, and the M entries, then feed the variables into a $Select statement or series of If and Else statements which represent boolean expressions of the information in the grey boxes. The result is of course the correct stage for that primary illustrated above. The evolution in the sophistication in the staging system is evident in the case of thyroid, where the Commission has broken the staging determinants further by including age, histology and grade into the equation. Although the staging for other cancers is based primarily by the anatomic extent of disease, it is not feasible to follow that pattern, since the morphological diagnosis and age play an important role in the prognosis of thyroid cancer and were therefore added to the staging system. After determining that the primary site is thyroid, the program first checks the grade, then morphology: depending on the findings it will then calculate age from the birthdate and date of diagnosis. These variables are fed into the Thyroid equation resulting in the appropriate stage. The dialog for the staging of Thyroid is shown in attachment #2.

The staging for breast (male included with female) is more detailed, and for gynecological tumors is even more so, with each particular stage grouping

(I,II,III,IV) divided into subgroupings (A,B,C) with the uterus having a total of eleven stages. Automatic staging becomes even more important as the number of possible stages increase and staging becomes more complex.

SUMMARY

These are just a few examples of the Automatic Staging routines in DHCP-Oncology Tumor Registry Package, Version 2. With the advent of the computerized registry, data collection, management and manipulation became efficient and effective. The addition of automatic staging of tumors helps eliminate coding errors and inconsistencies. The importance of these computerized programs are far reaching and have a tremendous potential. Virtually all clinicians associated with the diagnosis and treatment of cancer would benefit from an automatic staging module on the DHCP Clinician menu, just as many have found Dr. Fletcher's ENT staging extremely useful.

The next logical steps in this product's evolution would be to:

1) provide both clinical and pathological TNM staging,

2) include the First and Second staging manuals for old cases,

3) integrate of automatic staging with SEER extent of disease descriptions to fully allow automated start to finish staging -not a trivial task,.

4) integrate the various other DHCP-based products such as Anatomic Pathology, Radiology Hospital Summary/Operative notes and Imaging into a comprehensive package. The system would take the disjoined bits and pieces of information from all the packages, put it through logic and algorithms to produce a summary stage based on the extent of disease.

Automatic staging is but the beginning of an innovative and intelligent diagnostic tool which can only be improved upon.

References:

1) Fletcher M, McManus CD: A Data Management System for Head and Neck Oncology, Laryngoscope Dec 1987

2) Richie SR: Automatic Casefinding of Cancer Integration of Databases in the VA DHCP Automated Tumor Registry, MUMPS Computing 1993 Annual Proceedings.

3) Richie SR: Automated Tumor Registry for Oncology - A VA DHCP Mumps Application, SCAMC Proceedings Nov 1992

4) Richie SR: Hands on Demonstration of the VA-DHCP Automated Tumor Registry for Oncology, SCAMC Proc. Nov 1992

5) Richie SR: DHCP - Oncology Tumor Registry V2, MUMPS Computing 1992 Annual Proceedings.

6) Richie SR, Marciniak T File Manager and Statistics MUG Quarterly June 1991

7) Richie SR, Johnson Me File Manger Portals MUG Quarterly June 1991

Data Form for Cancer Staging

Patient identification
Name _____
Address _____
Hospital or clinic number _____
Age _____ Sex _____ Race _____

Institution identification
Hospital or clinic _____
Address _____

Oncology Record

Anatomic site of cancer _____
Histologic type _____
Grade (G) _____
Date of classification _____

Chronology of classification
[ ] Clinical (use all data prior to first treatment)
[ ] Pathologic (if definitively resected specimen available)

Clin	Path	DEFINITIONS
		Primary Tumor (T)
		All categories may be subdivided: (a) solitary; (b) multifocal—measure the largest for classification
[ ]	[ ]	TX Primary tumor cannot be assessed
[ ]	[ ]	T0 No evidence of primary tumor
[ ]	[ ]	T1 Tumor 1 cm or less in greatest dimension limited to the thyroid
[ ]	[ ]	T2 Tumor more than 1 cm but not more than 4 cm
[ ]	[ ]	T3 Tumor more than 4 cm in greatest dimension limited to the thyroid
[ ]	[ ]	T4 Tumor of any size extending beyond the thyroid capsule
		Lymph Node (N)
		Regional nodes are the cervical and upper mediastinal lymph nodes
[ ]	[ ]	NX Regional lymph nodes cannot be assessed
[ ]	[ ]	N0 No regional lymph node metastasis
[ ]	[ ]	N1 Regional lymph node metastasis
[ ]	[ ]	N1a Metastasis in ipsilateral cervical lymph nodes
[ ]	[ ]	N1b Metastasis in bilateral, midline, or contralateral cervical or mediastinal lymph nodes
		Distant Metastasis (M)
[ ]	[ ]	MX Presence of distant metastasis cannot be assessed
[ ]	[ ]	M0 No distant metastasis
[ ]	[ ]	M1 Distant metastasis

Clin	Path	Stage Grouping			
		Separate stage groupings are recommended for papillary and follicular, medullary, and undifferentiated			
		Papillary or Follicular			
		Under 45 Years			
[ ]	[ ]	Stage I	Any T, Any N, M0		
[ ]	[ ]	Stage II	Any T, Any N, M1		
		45 Years and Over			
[ ]	[ ]	Stage I	T1, N0, M0		
[ ]	[ ]	Stage II	T2, N0, M0		
			T3, N0, M0		
[ ]	[ ]	Stage III	T4, N0, M0		
			Any T, N1, M0		
[ ]	[ ]	Stage IV	Any T, Any N, M1		
		Medullary			
[ ]	[ ]	Stage I	T1	N0	M0
[ ]	[ ]	Stage II	T2	N0	M0
			T3	N0	M0
			T4	N0	M0
[ ]	[ ]	Stage III	Any T	N1	M0
[ ]	[ ]	Stage IV	Any T	Any N	M1
		Undifferentiated			
		All cases are Stage IV			
[ ]	[ ]	Stage IV	Any T	Any N	Any M

Staged by _____ M.D.
_____ Registrar
Date _____

(continued on next page)

American Joint Committee on Cancer—1992

```
                                                        Attachment #2
Patient name:    SVETERAN,JOHN Q              DOB: SEPT 3, 1910

Acc/Seq#    Primary Site          Tumor Status      Date DX
8-0249/00   THYROID               Evidence of CA    08/08/88

GRADE:// ?    I - Well Differentialed/Differentiated
             II - Moderately/Mod. well differentiated/Intermediate
            III - Poorly Differntiated
             IV - Undifferentiated, Anaplastic
GRADE: II
HISTOLOGY: FOLLICULAR ADENOCARCINOMA// ?    Note:date of Birth
                                                 date of diagnosis
   8020/30  CARCINOMA, UNDIFFERENTIATED          Grade
   8021/3   CARCINOMA, ANAPLASTIC NOS           Histology
   8050/3   PAPILLARY CARCINOMA NOS             T,N,M coding
   8260/3   PAPILLARY ADENOCARCINOMA            => AJCC Stage
   8290/0   OXYPHILIC ADENOMA
   8290/3   OXYPHILIC ADENOCARCINOMA
   8330/0   FOLLICULAR ADENOMA
   8330/3   FOLLICULAR ADENOCARCINOMA
   8331/3   FOLLICULAR ADENOCA, WELL DIFF
   8332/3   FOLLICULAR ADENOCA, TRABECULAR
   8340/3   PAPILLARY/FOLLICULAR ADENOCA
   8350/3   NONENCAPSULATED SCLEROSING CA
   8510/3   MEDULLARY CARCINOMA NOS
   8511/3   MEDULLARY CA, AMYLOID STROMA
T-code: // ?
     TX   Primary Tumor Cannot Be Assessed
     T0   No Evidence of Primary Tumor
     T1   1 cm OR LESS, LIMITED TO THYROID
     T1a  1 cm OR LESS, SOLITARY, LIMITED TO THYROID
     T1b  1 cm OR LESS, MULTIFOCAL, LIM TO THYROID
     T2   1 cm TO 4 cm, LIMITED TO THYROID
     T2a  1 cm TO 4 cm, SOLITARY, LIMITED TO THYROID
     T2b  1 cm TO 4 cm, MULTIFOCAL, LIM TO THYROID
     T3   OVER 4 cm, NON INVASIVE
     T3a  OVER 4 cm, SOLITARY, NON INVASIVE
     T3b  OVER 4 cm, MULTIFOCAL, NON INVASIVE
     T4   ANY SIZE EXTENDING BEYOND THYROID
     T4a  ANY SIZE, SOLITARY, EXTS BEYOND THYROID
     T4b  ANY SIZE, MULTIFOCAL, EXTS BEYOND THYROID
T-code: 2A//
N-code: // ?
     NX   Regional Lymph Nodes Cannot Be Assessed
     N0   No Regional Lymph Node Metastasis
     N1   REGIONAL LN METASTASIS
     N1a  METS IN IPSILAT CERVICAL LN
     N1b  BILAT/MIDLINE/CONTRALAT LN
N-code: 1B//
M-code: 0// ?
     MX   CAN'T ASSESS DISTANT METS
     M0   NO DISTANT METASTASIS
     M1   DISTANT METASTASIS
M-code: 0//

        Computed AJCC SUMMARY STAGE: III (T2A N1B M0)  3rd Ed.
```

Attachment #3

AUTOMATIC STAGING makes statistical analyisis such as this more accurate and uniform amoung facitlies using the VA-DHCP Package. This table was produced using the cross-tab routines in the package. The Rows, Selected Sites, represents a computed field taking which selects out important sites, and groups all remaining sites as 'other'. The stage groups is also computed, combining IIA and IIb etc into II.

ALL ANALYTIC CASES 1981 - 1991 BIRMINGHAM VAMC

SELECTED SITES:	0	I	II	III	Total
BLADDER	8	47	12	11	164
COLON	18	58	40	50	254
LEUK	0	0	0	0	42
LUNG,NSC	3	206	38	426	1614
LUNG,SC	0	12	2	46	146
LYMPH-H	0	4	5	3	21
LYMPH-NH	0	9	3	10	85
MELANOMA	13	19	6	4	76
ORAL CAV	4	20	20	11	147
OTHER	58	277	158	170	2023
PROSTATE	8	84	87	50	577
Total	112	736	371	781	5149

STAGE GROUPING-AJCC

SELECTED SITES:	IV	Unk/Uns	?	Total
BLADDER	14	35	37	164
COLON	64	13	11	254
LEUK	4	21	17	42
LUNG,NSC	185	338	418	1614
LUNG,SC	34	23	29	146
LYMPH-H	1	2	6	21
LYMPH-NH	13	27	23	85
MELANOMA	5	26	3	76
ORAL CAV	21	71	0	147
OTHER	296	852	212	2023
PROSTATE	105	126	117	577
Total	742	1534	873	5149