

INTERNATIONAL AGENCY FOR
RESEARCH ON CANCER (WHO)

INTERNATIONAL ASSOCIATION
OF CANCER REGISTRIES

CANCER REGISTRATION PRINCIPLES AND METHODS

IARC SCIENTIFIC PUBLICATIONS

N° 95

LYON 1991

がん登録の原理と方法

EDITORS: O.M. JENSEN, D.M. PARKIN,
R. MACLENNAN, C.S. MUIR and R.G. SKEET

厚生省がん研究助成金 5-3

「地域がん登録の精度向上と活用」に関する研究班 訳

1996. 3.

INTERNATIONAL AGENCY FOR
RESEARCH ON CANCER (WHO)

INTERNATIONAL ASSOCIATION
OF CANCER REGISTRIES

CANCER REGISTRATION PRINCIPLES AND METHODS

IARC SCIENTIFIC PUBLICATIONS

N° 95
LYON 1991

がん登録の原理と方法

EDITORS: O.M. JENSEN, D.M. PARKIN,
R. MACLENNAN, C.S. MUIR and R.G. SKEET

厚生省がん研究助成金 5-3

「地域がん登録の精度向上と活用」に関する研究班 訳

1996. 3.

IACRとの協約により、本書からの引用、転載をおことわりします。
引用は、原書から直接されるよう、また、転載については IACR の
Parkin 博士に許可を申請して下さるよう、お願いします。

主任研究者

本書翻訳の趣旨

わが国の地域がん登録は、1950年代後半から始まったが、その目的および技法はそれぞれの登録室で異なり、体系化されたものではなかった。1975年に発足した厚生省がん研究助成金による「地域がん登録」研究班は、「地域がん登録の手引き」を刊行し、1977年11月には「手引き、改訂版」を作成するとともに、「登録作業の標準方式」を刊行した。これらの内容は、1978年に国際がん研究所（IARC）と国際がん登録協会（IACR）とが共同して刊行した“Cancer Registration and its Techniques”（後述）の内容とも、ほぼ合致したものであった。

1983年に老人保健法が施行され、地域がん登録が行政的な基盤をもつに至った。そこでこれに対応すべく、研究班は、1986年に「手引き、第3版」を作成した。

その後、コンピュータ関連技術が進歩し、登録資料の活用範囲が拡大し、これに伴い活性化しつつある登録室もみられるが、開始したものの十分な成果をあげることが難しい県登録もあった。この様な状況の中で、本研究班では、地域登録の標準化をすすめることによって、改めてその体系化を促進すべく、「手引き、第4版」を作成することが必要と思われた。

一方、IARCとIACRは協同し、上述したごとく1978年に、“Cancer Registration and its Techniques” (eds.: Maclennan, R., Muir, C., Steinitz, R. and Winkler, A.)を刊行した。その後、世界各地でがん登録が拡がりを見せ、技術的にも進歩した。そこで13年を経過した1991年に、IARCとIACRとは再び協同して、“Cancer Registration, Principles and Methods” (eds.: Jensen, O.M., Parkin, D.M., Maclennan, R., Muir, C.S. and Skeet, R.G.)を刊行した。本書は、前書に比べ、電算機化された登録作業の方法、とくに質的精度管理、各種統計数値の計算方法、データ活用のための作表・作図方法、世界各地の登録室の活動例、などが含まれており、著者はいずれも、それぞれの分野の発展に努力してきた人々である。その内容は、わが国の地域がん登録室にとってその歴史の長短を問わず、益するところが多いと判断された。また、「手引き、第4版」作成にあたっては、参考すべき点が多いと考えた。

そこで、平成6年度研究班活動の中で、1994年5月に、翻訳サブグループ（津熊、味木、北川、陶山、藤田委由、花井）が集まり、担当する章をきめて作業に入った。同年8月に翻訳文をもちより、訳語の統一をはかり、9月に第一次訳文が出そろった。その後、各自担当分の見直しを行うとともに、藤田学、服部、井上が新たに参加し、津熊と共に訳文を推敲した。平成7年度に入って、さらに、馬淵と藤本が最終的に推敲し、藤本が図表を翻訳した。これらの全ての過程での再入力、編集、校正などの作業を、島、森、園部が担当した。

サブグループ発足から2年を経て、ようやく、印刷の運びとなった。私自身、初めての試みで、翻訳という作業を甘くみていたように思う。今日に至るまでの各作業に、長い時間とご努力を賜った方々にお詫び申し上げるとともに、衷心から御礼を申し上げる。

なお、IARCからは、がん登録関係者にのみ配布すること、日本語訳についての責任は訳者がもつこと、を条件に、翻訳の許可を得ている。

1996年3月 厚生省がん研究助成金 5-3

「地域がん登録の精度向上と活用」に関する研究班

主任研究者 花井 彩

翻訳グループ代表 津熊秀明

「がん登録の原理と方法」翻訳グループ

翻訳：津熊秀明(1)、味木和喜子(1)、花井 彩(1)、北川貴子(1)、

井上眞奈美(2)、陶山昭彦(3)、藤田委由(3)

推敲：津熊秀明(1)、藤田 学(5)、服部昌和(6)、

馬淵清彦(4)、花井 彩(1)

監修：藤本伊三郎(7)

編集、校正：島 尚子(1)、森 優子(7)、園部さやか(7)

(1)大阪府立成人病センター (2)愛知県がんセンター

(3)鳥取大医学部 (4)放射線影響研究所、広島

(5)市立敦賀病院 (6)福井県立勝山病院

(7)地域がん登録全国協議会

LIST OF AUTHORS

執筆者一覧 (ABC順、敬称略)

(所属機関が同じ場合は、2度目から住所を省略した。)

- Bieber, C. A. International Agency for Research on Cancer (IARC), 150 cours
Albert Thomas, 69372, Lyon Cédex 08, France.
- Boyle, P. IARC, Lyon, France.
- Clarke, E. A. Ontario Cancer Registry, Ontario Cancer Treatment and Research
Foundation, Toronto, Canada.
- Coleman, M. P. IARC, Lyon, France.
- Démaret, E. IARC, Lyon, France.
- Hakulinen, T. Finish Cancer Registry, Liisankatu 21B, 00170 Helsinki, Finland.
- van Holten, V. Cancer Statistics Branch, National Cancer Institute (NCI),
Bethesda, MD 20205, USA.
- Jensen, O. M. Danish Cancer Registry, Danish Cancer Society, Rosenvaengets
Hovedvej 35, PO Box 839, Copenhagen, Denmark.
- Kreiger, N. Ontario Cancer Registry, Toronto, Canada.
- MacLennan, R. Queensland Institute of Medical Research, Bramston Terrace,
Herston, Brisbane, QLD 4006, Australia.
- Marrett, L. D. Ontario Cancer Registry, Toronto, Canada.
- Muir, C. S. IARC, Lyon, France.
- Parkin, D. M. IARC, Lyon, France.
- Percy, C. NCI, National Institutes of Health, Bethesda, MD 20892-4200, USA.
- Powell, J. Birmingham & West Midlands Regional Cancer Registry, Queen
Elizabeth Medical Centre, Birmingham B15 2TH, UK.

Shanmugaratnam, K. Department of Pathology, National University of Singapore,
National University Hospital, Lower Kent Ridge Road, Singapore,
0511, Republic of Singapore.

Skeet, R. G. Herefordshire Health Authority, Victoria House, Hereford HR4 0AN,
UK.

Storm, H. H. Danish Cancer registry, Copenhagen, Denmark.

Wagner, G. German Cancer Research Centre, In Neuenheimer Feld 280, 6900
Heidelberg 1, Federal Republic of German.

Whelan, S. IARC, Lyon, France.

Young, J. L. California Tumor Registry, 1812 14th Street, Suite 200,
Sacramento, CA 95814, USA.

CONTENTS

目次

Chapter 1.	Introduction	K. Shanmugaratnam	
第1章	序論	津熊秀明、服部昌和 訳	--- 1
Chapter 2.	History of cancer registration	G. Wagner	
第2章	がん登録の歴史	津熊秀明、服部昌和 訳	--- 3
Chapter 3.	Purposes and uses of cancer registration	O. M. Jensen and H. H. Storm	
第3章	がん登録の目的と理由	津熊秀明、服部昌和 訳	--- 7
Chapter 4.	Planning a cancer registry	O. M. Jensen and S. Whelan	
第4章	がん登録設置計画	味木和喜子、花井 彩 訳	--- 23
Chapter 5.	Data sources and reporting	J. Powell	
第5章	データ源と届出票作成	味木和喜子、馬淵清彦 訳	--- 30
Chapter 6.	Items of patient information which may be collected by registries		
		R. MacLennan	
第6章	がん登録室において収集する患者情報項目	井上眞奈美、花井 彩 訳	--- 45
Chapter 7.	Classification and coding of neoplasms	C. S. Muir and C. Percy	
第7章	新生物の分類とコードディング	北川貴子、津熊秀明 訳	--- 68
Chapter 8.	Manual and computerized cancer registries	R. G. Skeet	
第8章	手作業及び電算機によるがん登録	津熊秀明、藤田 学 訳	--- 87
Chapter 9.	Quality and quality control	R. G. Skeet	
第9章	質的精度とその精度管理	花井 彩、藤田 学 訳	--- 108
Chapter 10.	Reporting of results	O. M. Jensen and H. H. Storm	
第10章	集計結果報告	味木和喜子、花井 彩 訳	--- 115
Chapter 11.	Statistical methods for registries	P. Boyle and D. M. Parkin	
第11章	がん登録における統計学的手法	味木和喜子、花井 彩 訳	--- 131
Chapter 12.	Analysis of survival	D. M. Parkin and T. Hakulinen	
第12章	生存率解析	味木和喜子、花井 彩 訳	--- 162

Chapter 13.	The hospital-based cancer registry	J. L. Young	
第13章	院内がん登録	北川貴子、津熊秀明、陶山昭彦、服部昌和 訳	--- 179
Chapter 14.	Cancer registration in developing countries	D. M. Parkin and L. D. Sanghvi	
		(省略)	
Chapter 15.	Cancer registration : legal aspects and confidentiality	C. S. Muir and E. Demaret	
第15章	がん登録：法的側面と情報保護	花井 彩 訳	--- 188
Appendix 1.	United Nations Standard Country Codes		
付録 1	国際連合標準国別コード	藤田委由、津熊秀明 訳	--- 198
Appendix 2.	Editing for consistency of data items	V. Van Holten	
付録 2	データ項目の整合性に関する編集作業	藤田委由、津熊秀明 訳	--- 202
Appendix 3.			
付録 3			
(a)	The Danish Cancer Registry, a self-reporting national cancer registration system with elements of active data collection	H. H. Storm	
	デンマークがん登録—能動的データ収集の要素を兼ね備えた 篤志届出による全国がん登録システム—	藤田委由、津熊秀明 訳	--- 211
(b)	The Thames Cancer Registry	R. G. Skeet	
	テムズがん登録室	藤田委由、津熊秀明 訳	--- 227
(c)	Cancer registration in Ontario	E. A. Clarke, L. D. Marrett and N. Kreiger	
	オンタリオ州におけるがん登録	味木和喜子、花井 彩 訳	--- 236
(d)	The Department of Health-Rizal Cancer Registry	A. V. Laudico and D. Esteban	
		(省略)	
Appendix 4.	CANREG : Cancer registration software for microcomputers	M. P. Coleman and C. A. Bieber	
付録 4	CANREG : マイクロコンピュータ用がん登録ソフトウェア	陶山昭彦、津熊秀明 訳	--- 248

第1章 序 論

K. Shanmugaratnam

がん登録は、がんの制圧において中心的な役割を担っている。がん登録の第一義的機能は、特定の人口集団に発生する全がん症例のファイル、もしくは登録簿を維持することである。ファイルもしくは登録簿には、種々のデータベースから、持続的かつ系統的に、集められるがん患者の個人的事項と、がんの臨床的、病理学的特性が記録される。がん登録は、そうしたデータを、定期的に解析、解釈し、居住人口の種々の区分における特定のがんの罹患及び特性について、さらには、罹患数の経時的変動に関する情報を提供する。こうした情報は、単にがんの決定要因に関する疫学的研究の為だけではなく、がんの予防と診断、治療の為のヘルスサービスの計画と評価の為の主要情報源でもある。

がん登録は、種々の発がん物質にさらされる職種グループや、コホート集団のモニタリングのためにも使われるし、臨床及び疫学調査のための調査対象の便利な供給源として活用し得る。院内登録は、患者のフォローに関する臨床医の援助、又治療成績に関する統計データを提供することにより、がん患者のケアにおいて重要な支援的役割を果たす。

がん登録の価値は、データの質と、研究及び保健医療計画において、データが利用される程度とに依存している。がん患者の登録は、できるだけ完全であるべきであるが、情報に関する機密を確保するために制定された法律や規則によって、がん登録実施に重大な支障を被っている所もある。これは、死亡記録に記載された死亡者の個人識別情報へのアクセスを含む、異なるデータファイル間の照合が阻害されるからである。その地域の状況に応じた機密に関する規約の下で運営することにより、がん登録がこうした情報にアクセスできることが望まれる。

現在活動しており、かつ、その成績が IARC シリーズの「5大陸のがん罹患」の中に掲載されている登録室の多くは、ヨーロッパおよび北アメリカに存在する。アジア、アフリカ、及び南アメリカの開発途上国で、より多くのがん登録を実施することが緊急の課題である。こうした地域では、がんが既に主要な健康問題の一つとして認識されており、更に、感染症の制圧及び寿命の延長と共に、がんの重要性が高まると考えられる。個々の登録室で収集されるデータは、地域の需要と利用できる情報とにより様々である。しかし、各項目の用語規約及び定義は、がんデータの国際的な比較性を高めるために、すべての登録室で同一であるべきである。又、すべての登録室が収集に努めるべき国際的に合意された基本データ項目がある。本書の目的の一つは、こうしたデータ収集の統一性を高めることにある。

本書は、IARC（国際がん研究機関）と IACR（国際がん登録学会）との協力の結果ででき上が

った。本書の目的は、がん登録のあらゆる面におけるガイドラインを提供することにある。本書は以前の出版物「がん登録とその技法 (Cancer Registration and its Techniques, IARC Scientific Publications No.21)」に取って代わるものであるが、元の書とはいくつかの面で異なっている。本書では、多数の著者が各々の様式で記述しており、がん登録が電算機を基本としたものになるであろうとの、全体としての想定がある。

がん登録の利用法については本書で、より詳細に記載されている(第3章)が、がん関連の保健サービスを計画し、評価する際のがん登録の重要性は、IARC モノグラフ「がん制圧におけるがん登録の役割 (The Role of the Registry in Cancer Control, IARC Scientific Publications No.66)」で、より詳細に扱われている。がん制圧の疫学的、公衆衛生的局面に主要な関心のある地域ががん登録に力点がある。登録が推奨されるデータ項目は、最小限にとどめ、情報の量よりもむしろ質を強調した。なぜなら、項目は、もし必要なら地域のニーズに合わせて、拡張できるからである。第13章で述べているように、院内がん登録の運営は、より患者のケア、臨床研究、及び病院管理に結びついている。院内がん登録は、地域がん登録が直ちには構築できない国々において、将来の構築のための核として役立つ。

本書で述べるがん登録運営の方法が、とりわけがんの罹患と特性について、未だ十分には記述されていない国々において、より多くのがん登録の構築を促進し、そして、がん登録の全ての局面において、統一的な方法の採用により収集されたデータの有用性を最大にする上で、役立つことを希望している。

(翻訳：津熊秀明、服部昌和)

第2章 がん登録の歴史

G. Wagner

草創期

がんを患っている人々の登録は、多くの曲折と行き詰まりを伴いながら、ゆっくりとしたスピードで発展してきた。この章では、がん登録の歴史を簡潔に要約する。詳しい総説には、Clemmesen (1965) と Wagner (1985) を参照のこと。

失敗に終わったが、1728年 Londonで、最初がんに関する人口調査が行われた。信頼できる、比較可能な死亡と罹患に関する統計を作ろうとした試みは、今世紀の初頭までは失敗に終わり、現実的には殆ど把握されなかった (Kennaway, 1950)。1900年頃、England や、とくにドイツにおいて、病因論の研究に欠くことのできない基礎として、地域でのがんの広がりに関する質の高い統計調査を要求する声があがった。Katz (1899) は、ハンブルクでがんに関する概括的調査を要求した。そして、1900年には、ドイツにおいて治療中の全がん患者を登録しようとする試みがあった。1900年10月15日現在のがんの有病率を記録するために、プロシア文化大臣を通じて、ドイツの全内科医にアンケート調査票を送付した (がん問題に関する委員会、1901)。

この試みは、1902-08年にオランダ、スペイン、ポルトガル、ハンガリー、スウェーデン、デンマーク、アイスランドで繰り返された。ドイツでの調査報告の中で、アンケート調査票に記入し、返送した者は、アンケートを送付したものの内、半分を僅かにこえるに止まったとの記載があった (von Leyden ら、1902)。ハイデルベルグ (1904) やバーデン (1906) で、国レベルのがん罹患統計を得ようとした同様の試みと同じく、調査は失敗とみなされた (Hecht, 1902)。1905年のロンドン大英帝国がん研究財団の報告書の中で、Bashford and Murray (1905) は、ドイツでの多くの努力にもかかわらず、存在していたであろうがん問題の多くを未解明のまま残したことを見出して、がんについての全国実態調査を行うことに反対の意見を述べた。これらの調査のほとんどは、医師の協力が不十分であったゆえに、Woodは、米国においては、がんが届出の必要な病気とし、全がん患者の義務的登録が実施されるべきであると、指摘した。一方、マサチューセッツ州で1927年のがん登録がパイロットとしてスタートしたが、約1/3しかがん患者が報告されなかったことから、失敗と見なされた (Hoffman, 1930)。

がん罹患統計を作成する目的で、1937年に Mecklenburg (旧東ドイツ) で、がん患者の継続的記録が始まった (Lasch, 1940)。ここでは、方法論の進歩があった。すなわち、名前による報告が、初めて重複登録を避け、個人個人の子後を決定的にすることを可能にした。全ての開業医、病院、及び病理検査施設が登録カードや様式を受け取り、そこにがん患者の情報を

記入し、2週毎に Rostock (旧東ドイツ) の統計事務所に送らねばならなかった。そこで報告書がチェックされ、索引カードに入れられる。報告のない場合は、毎日督促者によって、電話で報告が求められる。1937-38年のカバー率は、100,000人当たり約200人の新発生患者で、これで示される通り、この登録制度は随分うまく稼働したように見える (Wagner, 1985)。この好ましい経験に引き続き、同様の調査が、1939年に Saxony-Anhalt (訳者注: 場所不明)、ザールランド、ウイーンで始まった。しかしながら、政治的理由で、まもなく中止せざるを得なかった。

ほぼ同じ頃、米国でがん罹患特別調査で、がん罹患データを収集する試みが行われた。1937-38年に、10大都市での1年間の全がん患者が記録された。この National Cancer Survey は、1947-48年と1969-71年にも繰り返された。米国とヨーロッパにおけるこれら初期の調査の唯一の目的は、種々のがんの罹患率、死亡率、及び有病率に関する蓄積であった。これらの調査でカバーされたがん患者の予後は不明であった。それ故、米国での第3回 National Cancer Survey は、この種の最終のものと、予め決められた。なぜなら、継続的な登録が、最終結果に関する調査のためには優れていると考えられたからである。

現代の発展

現代のがん登録の最も古い例は、ハンブルクの登録である。本登録は、がんの制圧は、単に医学的、科学的側面のみを含むのではなく、公衆衛生や経済面を含んでいるとの考えを持って、始まった。1926年に、がん患者の後療法のための組織が私的に設立された。1929年以来、この施設は、ハンブルク公衆衛生部の、フォローアップ患者ケアサービスとして、公の地位を獲得した (Bierich, 1931; Sieveking, 1930, 1933, 1935, 1940)。3人の看護婦がハンブルクの病院と開業医を定期的に訪問し、新発生患者の名前を記録し、保健部にある中央登録室にデータを転送した。こうして出来たカードは、公的な死亡診断書と、週に1回比較され、ハンブルクがん登録の基礎となった (Keding, 1973)。

疫学的、生態学的目的を有する地域がん登録が、米国において1935年に始まった。この年に、コネチカット州保健部に、「がんに関する研究、すなわち、がんの予防と治療、がんによる死亡に関する研究を行い、がんによる死亡の減少をもたらすのを援助するような行動をとるために」がん研究部門が創設された。コネチカット州腫瘍登録は、1941年に州全域の登録を開始し、1935年にまで遡って、がん症例を登録した (Griswold et al, 1955; Connelly et al, 1968)。1940年代の初期に、米国とカナダにおいて、地域がん登録がさらに設立された (Stocks, 1959; Barclay, 1976)。

デンマークがん登録は、デンマーク対がん協会の援助の下で、1942年に設立され、一国を

カバーする最も古いがん登録である。デンマーク医師会の後援をも得て、がん患者が開業医によって自主的に届け出られ、一方、National Board of Health は、死亡診断書と全死亡データの閲覧許可を与え、援助している。登録の仕事は、以下の a~c のための基礎的なことからを提供するためのデータ集積として枠組まれた。すなわち、(a)患者の個々のフォローアップのため；(b) 治療結果の正確な予測についての観点を備えた信頼できる罹患統計のため；そして(c) 悪性新生物の罹患率の変化、すなわち、地理的、職業的変動等と共に、年次動向の正確な評価のため (Clemmesen、1965)。1940年代の中期から、表1に掲載したいくつかの国で、がん登録が開始された。

がん登録の世界的な規模での設立に向けての最も重要な機運は、恐らく1946年にデンマークがん登録室長、Clemmesen 博士の先導により、コペンハーゲンで開催されたカンファレンスに由来している (Schinz、1946)。がん制圧分野の国際的専門家12人からなるグループが、WHO の Interim Commission にがん登録の世界規模での設立を訴えた。彼らは、

- (a) 出来るだけ多くの異なった国からがん患者のデータを収集すれば、大きな利益が得られるであろう；
- (b) そうしたデータは、比較が可能ないように合意された計画に基づいて記録されるべきである；
- (c) 各国には、データの記録と収集について調整する中央登録室 (central registry) を置くべきである；
- (d) 各国で得たデータと統計値を対比させるために、国際機関を持つべきである。

といったことを提言した。

表1. 1955年以前に開設した地域がん登録室

国名 (地域名)	開設年	届出の法律
ドイツ (ハンブルグ)	1929	なし
米 国 (ニューヨーク州)	1940	あり
" (コネチカット)	1941	あり (1971年以後)
デンマーク	1942	あり (1987年以後)
カナダ (サスカチワン)	1944	あり
英 国 (ロンドン南西部)	1945	なし
" (リバプール)	1948	なし
ニュージーランド	1948	あり
カナダ (マニトバ)	1950	なし
ユーゴスラビア (スロベニア)	1950	あり
カナダ (アルベルタ)	1951	あり
米 国 (エルパソ)	1951	なし
ハンガリー (3地域)	1952	あり
ノルウェー	1952	あり
ソ 連	1953	あり
東ドイツ	1953	あり
フィンランド	1953	あり (1961年以後)
アイスランド	1954	なし

それから4年後に WHOは、がん患者の登録と統計の提示に関する小委員会を設立し、当該委員会ががん登録の確立に向けて勧告を出した (Stocks, 1959)。1959年に国際対がん連合 (UICC) によって準備されたがんの地理病理学、人口統計学に関する国際シンポジウムの際に、これはまた別の記念碑になったが、ある限られた地域で、がんの新発患者を数え上げることの必要性が強調された (Clemmesen, 1951)。シンポジウムでの勧告に基づいて、UICCは地理病理に関する委員会を設立した。1965年には、国際がん研究所 (IARC) が、WHOの特別がん研究センターとして設立された。

こうした進歩の自然の結果として、国際がん登録学会 (IACR) が1966年に東京で結成された。IACRは、「がん罹患データの収集と分析に関する、そして、地域集団におけるがん治療の最終成績に関する」がん登録の会員制の組織として、役割をはたしている。IACRは、IARCと密接に協力している。

がん登録の歴史的発展は、このように鮮明にたどることが出来る。約200の地域がん登録が、世界の種々の地域に存在している (Coleman & Wahrendorf, 1989)。さらに、特定の年齢グループもしくは部位 (例えば、ドイツの Mainz、英国のオックスフォード、及びオーストラリアの小児がん；フランスのDijon の消化管がん) の登録のみを実施している約34の登録がある。また、多くの病院が、院内がん登録を構築している (第13章参照)。

(翻訳：津熊秀明、服部昌和)

第3章 がん登録の目的と利用

O. M. Jensen and H. H. Storm

がん登録は、がんの制圧に関する合理的プログラムの中の欠くことの出来ない部分である (Muir et al., 1985)。がん登録データは、病因論に関する研究から、第一次及び第二次予防、ヘルスケア計画と患者のケアに至るまで、がん制圧に関する種々の広い分野で、利用することが出来る。それ故、個人と社会の双方に利益をもたらす。殆どのがん登録は、データ利用のための基礎を提供すること以上の義務を負っているわけではないが、がん登録は、収集した情報を使って、重要な研究計画を進展させ支援する能力を有している。

がん登録の主要な目的は、全がん症例の情報を集め、分類することにより、ある一定の人口におけるがんの発生に関する統計を作成し、その地域社会におけるがんの影響を評価し、制御するための枠組みを提供することである。この目的は、今から50年前に、初めてがん登録が設立され (第2章)、がん登録が、「研究の基礎を打ち立てる事だけに義務を負った」時と同じく、正当である (Clemmesen, 1965)。

しかしながら、がん症例の情報を収集し、がんの統計を作成することは、収集したデータが利用される場合のみ正当化される。がん登録情報は、多数の分野で用い得るが、もしも比較性が長く維持されておれば、データの価値が増大する。この章では、疫学研究、ヘルスケア計画とモニタリング、そしてその他の分野におけるがん登録データの利用の例を提示する。

地方の事情と興味により、がん登録の力点は、ところにより異なる。一般的な言い方をすれば、データは、長い期間にわたって蓄積されれば、益々多目的に利用可能になる。

疫学研究

がん疫学研究者は、がんの決定要因の探索のために、人間集団におけるがんの分布に関する知識を活用する。がん登録は、非致死例を含むがんの分布に関する情報を保持しているのであるから、明らかに、疫学のための重要な基盤を提供する。しかしながら、罹患数を算定することに加えて、一定の集団に由来するがん患者の記録を収集することにより、個人々人におけるがんの研究を深くさせ、一方で、臨床研究で見出されるセレクションバイアスを最小限にする。以下の記述では、記述疫学のためのがん登録の活用と、分析疫学のための活用とを区別している。しかしながら、疫学のこれら2つの側面は、相補的で、かつ、しばしばオーバーラップすることを、強調しておかなければならない。

表1. デンマークにおけるがん統計、1982年、特定部位
(1985年デンマークがん登録室の成績)

男

部位	罹患数	罹患率 人口10万対		有病数	有病率 人口10万対
		粗率	年齢調整率*		
全部位	11,533	457.4	297.6	49,471	1962.0
口腔、咽頭	334	13.2	9.2	3,128	124.1
胃	547	21.7	13.1	1,037	41.1
肺	2209	87.6	56.5	2,401	95.2
乳房	27	1.1	0.7	125	5.0
子宮頸	—	—	—	—	—
睪丸	230	9.1	8.2	2,895	114.8
皮膚黒色種	204	8.1	6.3	1,406	55.8
ホジキン病	74	2.9	2.4	732	29.0

女

全部位	11,723	451.9	270.2	80,744	3112.5
口腔、咽頭	144	5.6	3.1	1,218	47.0
胃	353	13.6	6.2	725	27.9
肺	779	30.0	18.5	901	34.7
乳房	2,469	95.2	63.8	21,318	821.8
子宮頸	638	24.6	18.9	12,014	463.1
睪丸	—	—	—	—	—
皮膚黒色種	273	10.5	7.8	2,973	114.6
ホジキン病	52	2.0	1.7	521	20.1

* 世界標準人口

記述疫学研究

がん登録による一定集団でのがん患者数の把握は、新発生患者数と罹患率計算の点から、がん問題の大きさの評価を可能にする。がん登録からでてくる統計の形式は、国際比較の重要性を考えつつ、現地のニーズと興味に適合させるべきである（例と計算については第10-12章を参照せよ）。率を計算できるかどうかは、分母である人口数を利用できるかどうかにかかっている。当然、がん患者情報は、統計情報部によって作成される人口集団に合致するように、収集、分類されなければならない。基本的な記述統計は、主として腫瘍の解剖学的部位分類に従う診断区分（第7章）別に、作成、提示されるべきである。殆んどどの部位で、がんは頻度が少なく、それ故、数のランダムな変動を最小にするため、患者を数年間にわたって集積する必要があるかも知れない（第10章、11章）。

罹患に関する数値に加えて、がんの有病率統計は、地域におけるがん発生に関する基礎的情報を、完全なものとする。有病率に関する統計は、罹患と生存に関するデータから、推計することもできる（MacMahon & Pugh, 1972; Hakama et al., 1975）。しかし、がん登録が長年にわたって機能していて、登録の設立以前にがんと診断された全ての患者が死亡している場合には、有病者を、登録ファイルから単純に数え上げることが出来る。ただし、もちろん、登録が死亡と転出に関する情報を受け取っている場合に限られる（デンマークがん登録、1985）。

表1に、がん登録統計の基本例を示す。登録成績の報告に関する詳細な記述は、第10章にある。

種々の集団におけるがん発生の比較は、病因に関する手がかりを提供する場合がある。そして、罹患（と死亡）の相違を示すことが、多くのがんの環境要因の存在の認識に大きく貢献し、従って、予防の可能性を示すこととなった（Higginson & Muir, 1979; Doll et al., 1981）。性、年齢別の統計は、大変異なったパターンを示し、部位間で変動がある（図1）。がん罹患に関するそうした基本的特徴は、必ずしも容易に理解され、説明されないかも知れない。しかし、それらは疫学者の興味を引き起こし、病因に関する仮説を生み出すのに、当然有益である。

がん登録の、がん罹患の国際間の相違に関する知識への貢献は、しばしばがん患者登録の目的として見すごされているが、重要である。系統的な比較は、モノグラフ「5大陸のがん罹患」（Doll et al., 1966; Muir et al., 1987; Waterhouse et al., 1970, 1976, 1982）として出版されている。がん罹患に関するそうした地理比較に由来する疫学的考えは、可能性のあるリスク要因の統計と関連させることにより、より強く促進されるかも知れない（例、Armstrong & Doll, 1975）。がんの国際間のパターンは、例えば、子宮頸がんの罹患率に5-6倍の差のあるグリーンランドとデンマークにおける human papilloma virus 感染の比較

表2. がんの部位別罹患率の世界各地における変動
(罹患数が10以下の場合を除いた。データはC I - 5, 第7巻による)

Site	ICD-9	Males			Females		
		Highest	Lowest	Ratio of highest to lowest	Highest	Lowest	Ratio of highest to lowest
Lip	(140)	Canada, Newfoundland 15.1	Japan, Osaka 0.1	151.0	Australia, South 1.6	U.K. England & Wales 0.1	16.0
Oral cavity	(143-14)	France, Bas-Rhin 13.5	Japan, Miyagi 0.5	27.0	India, Bangalore 15.7	Japan, Miyagi 0.2	78.5
Nasopharynx	(147)	Hong Kong 30.0	UK, South Wales 0.3	100.0	Hong Kong 12.9	USA, Iowa 0.1	129.0
Oesophagus	(150)	France, Calvados 29.9	Romania, County Cluj 1.2	24.9	India, Poona 12.4	Czechoslovakia, Slovakia 0.3	41.3
Stomach	(151)	Japan, Nagasaki 82.0	Kuwait, Kuwaitis 3.7	22.2	Japan, Nagasaki 36.1	USA, Iowa 3.0	12.0
Colon	(153)	USA, Connecticut, whites 34.1	India, Madras 1.8	18.9	USA, Detroit, blacks 29.0	India, Nagpur 1.8	16.1
Rectum	(154)	FR Germany, Saarland 21.5	Kuwait, Kuwaitis 3.0	7.2	FR Germany, Saarland 13.2	India, Madras 1.3	10.2
Liver	(155)	China, Shanghai 34.4	Canada, Nova Scotia 0.7	49.1	China, Shanghai 11.6	Australia, N.S. Wales 0.4	29.0
Pancreas	(157)	USA, Los Angeles, Koreans 16.4	India, Madras 0.9	18.2	USA, Alameda, blacks 9.4	India, Bombay 1.3	7.2
Larynx	(161)	Brazil, São Paulo 17.8	Japan, Miyagi 2.2	8.1	USA, Connecticut: Black 2.7	Japan, Miyagi 0.2	13.5
Lung	(162)	USA, New Orleans, blacks 111.0	India, Madras 5.8	19.0	New Zealand, Maoris 68.1	India, Madras 1.2	56.8
Melanoma	(172)	Australia, Queensland 30.9	Japan, Osaka 0.2	154.5	Australia, Queensland 28.5	India, Bombay 0.2	142.5
Other skin	(173)	Australia, Tasmania 167.2	India, Madras 0.9	185.8	Australia, Tasmania 89.3	Switzerland, Zurich 0.6	148.8

Breast	(175/174)	Brazil, Recife 3.4	Finland 0.2	17.0	Hawaii, Hawaiian 93.9	Israel, non-Jews 14.0	6.7
Cervix uteri	(180)	—	—	—	Brazil, Recife 83.2	Israel: non-Jews 3.0	27.7
Corpus uteri	(182)	—	—	—	USA, San Francisco Bay Area, whites 25.7	India, Nagpur 1.2	21.4
Ovary, etc.	(183)	—	—	—	NZ, Pacific Polyn. Isl. 25.8	Kuwait, Kuwaitis 3.3	7.8
Prostate	(185)	USA, Atlanta: blacks 91.2	China, Tianjin 1.3	70.2	—	—	—
Testis	(186)	Switzerland, Basle 8.3	China, Tianjin 0.6	13.8	—	—	—
Penis, etc.	(187)	Brazil, Recife 8.3	Israel: All Jews 0.2	41.5	—	—	—
Bladder	(188)	Switzerland, Basle 27.8	India, Nagpur 1.7	16.4	Kuwait, non-Kuwaitis 8.5	India, Poona 0.8	10.6
Kidney, etc.	(189)	Canada, NWT and Yukon 15.0	India, Poona 0.7	21.4	Iceland 7.6	India, Poona 0.6	12.7
Brain	(191/192)	NZ, Pacific Polyn. Isl. 9.7	India, Nagpur 1.1	8.8	Israel, born Israel 10.8	India, Madras 0.8	13.5
Thyroid	(193)	Hawaii, Chinese 8.8	Poland, Warsaw City 0.4	22.0	Hawaii, Filipinos 18.2	India, Nagpur 1.0	18.2
Lympho- sarcoma	(200)	Switzerland, Basle 9.2	France, Calvados 0.9	10.2	Australia, Cap. Territ. 7.2	Japan, Miyagi 0.4	18.0
Hodgkin's disease	(201)	Canada, Quebec 4.8	Japan, Miyagi 0.5	9.6	Switzerland, Neuchatel 3.9	Japan, Osaka 0.3	13.0
Multiple myeloma	(203)	USA, Alameda, blacks 8.8	Philippines, Rizal 0.4	22.0	USA, Connecticut, blacks 7.4	China, Shanghai 0.4	18.5
Leukaemia	(204-8)	Canada, Ontario 11.6	India, Nagpur 2.2	5.3	Pacific Polyn. Isl. 10.3	India, Madras 1.1	9.4

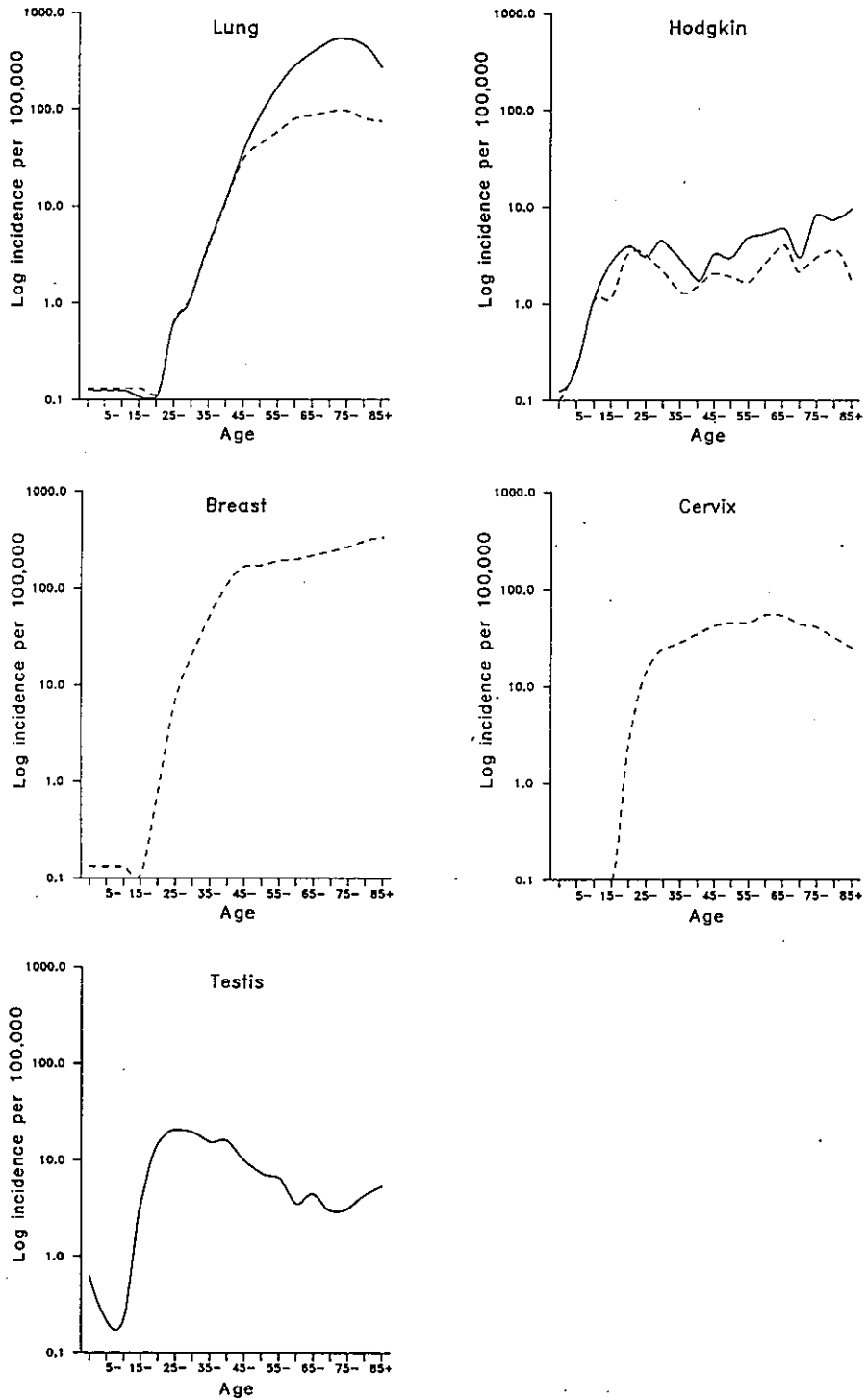


図1. がんの部位別、年齢階級別罹患率曲線のパターン（デンマーク、1978-82年の成績）

表3. いくつかの国でのがんの罹患率の市郡間の比の比較、性別、部位別、1980年
(データは Muirら「5大陸のがん罹患. 第5巻」1987年による)

Tumour site	France, Doubs		Norway		Japan, Miyagi Prefecture	
	M	F	M	F	M	F
Oesophagus	0.8	0.6	1.7	1.3	1.0	1.0
Stomach	1.2	1.1	1.0	1.0	1.0	1.0
Colon	1.0	1.2	1.2	1.1	1.3	1.4
Rectum	1.2	0.9	1.2	1.2	1.1	1.1
Larynx	1.3	0.8	1.6	2.0	1.1	0.7
Lung	1.4	1.7	1.6	1.8	1.1	1.1
Melanoma of skin	1.1	0.9	1.3	1.3	1.4	2.0
Breast	—	1.1	—	1.2	—	1.3
Cervix uteri	—	1.0	—	1.3	—	1.5
Testis	0.9	—	1.2	—	1.1	—
Bladder	1.1	1.2	1.3	1.6	1.1	1.8
Hodgkin's disease	0.7	0.7	1.0	1.1	0.4	1.0

表4. 移民のがん罹患に及ぼす影響—宮城県、サンフランシスコ湾岸地区の日本人と白人との3群のがん罹患率の比較 (データは Muirら「5大陸のがん罹患. 第5巻」1987年による)

Tumour site	Males			Females		
	Miyagi	Japanese (Bay Area)	White (Bay Area)	Miyagi	Japanese (Bay Area)	White (Bay Area)
Stomach	79.6	24.3	10.4	36.0	10.8	4.8
Colon	9.8	29.8	30.6	9.4	20.8	23.7
Rectum	9.9	13.6	15.4	7.4	12.4	11.0
Lung	29.6	33.0	65.8	8.7	12.1	33.3
Breast	—	—	—	22.0	48.9	87.0
Cervix uteri	—	—	—	10.0	5.9 ^a	8.9
Corpus uteri	—	—	—	2.8	19.6	25.7
Ovary	—	—	—	4.2	8.8	12.9
Prostate	6.3	16.5	50.0	—	—	—

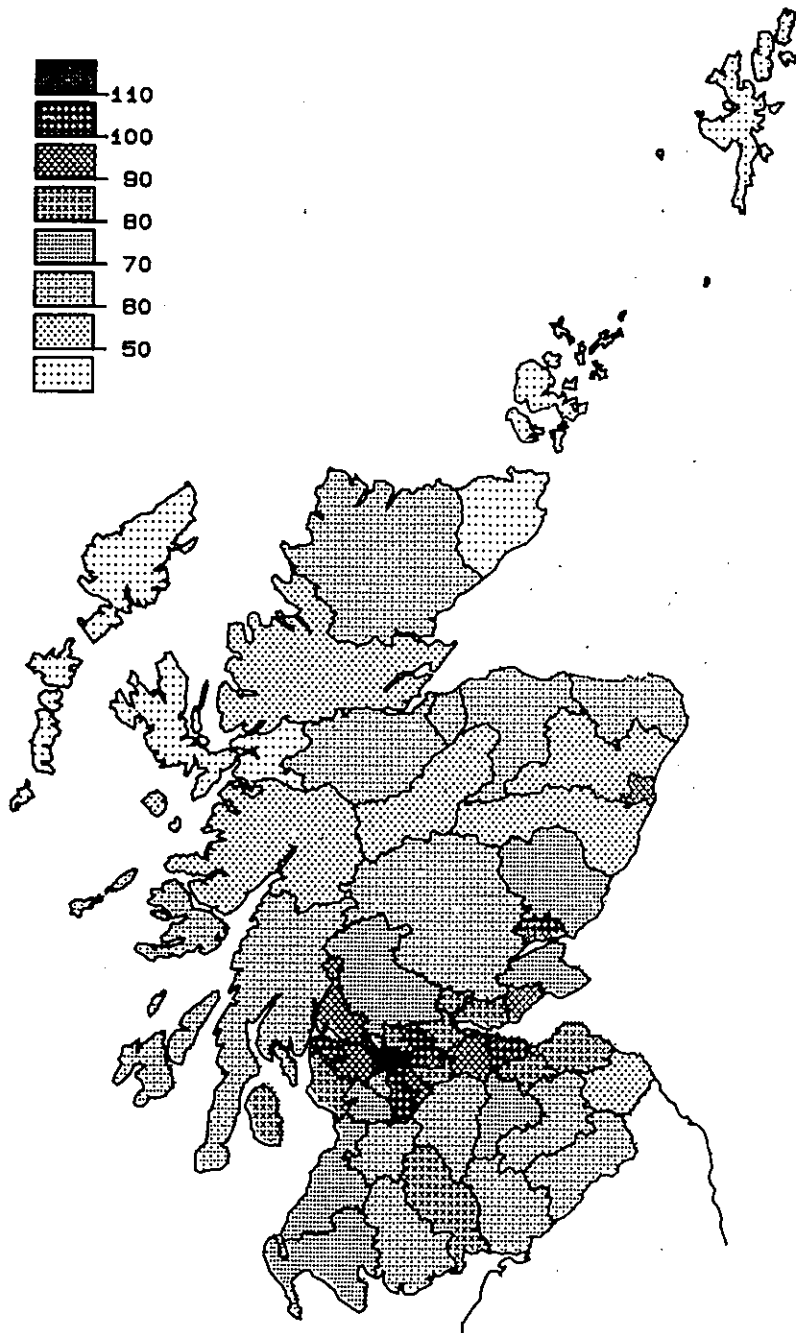


図2. 肺がん罹患地図—男、スコットランド、1975-80年
 (Kemp ら 1985年による) 年齢調整罹患率を%の形で比較している

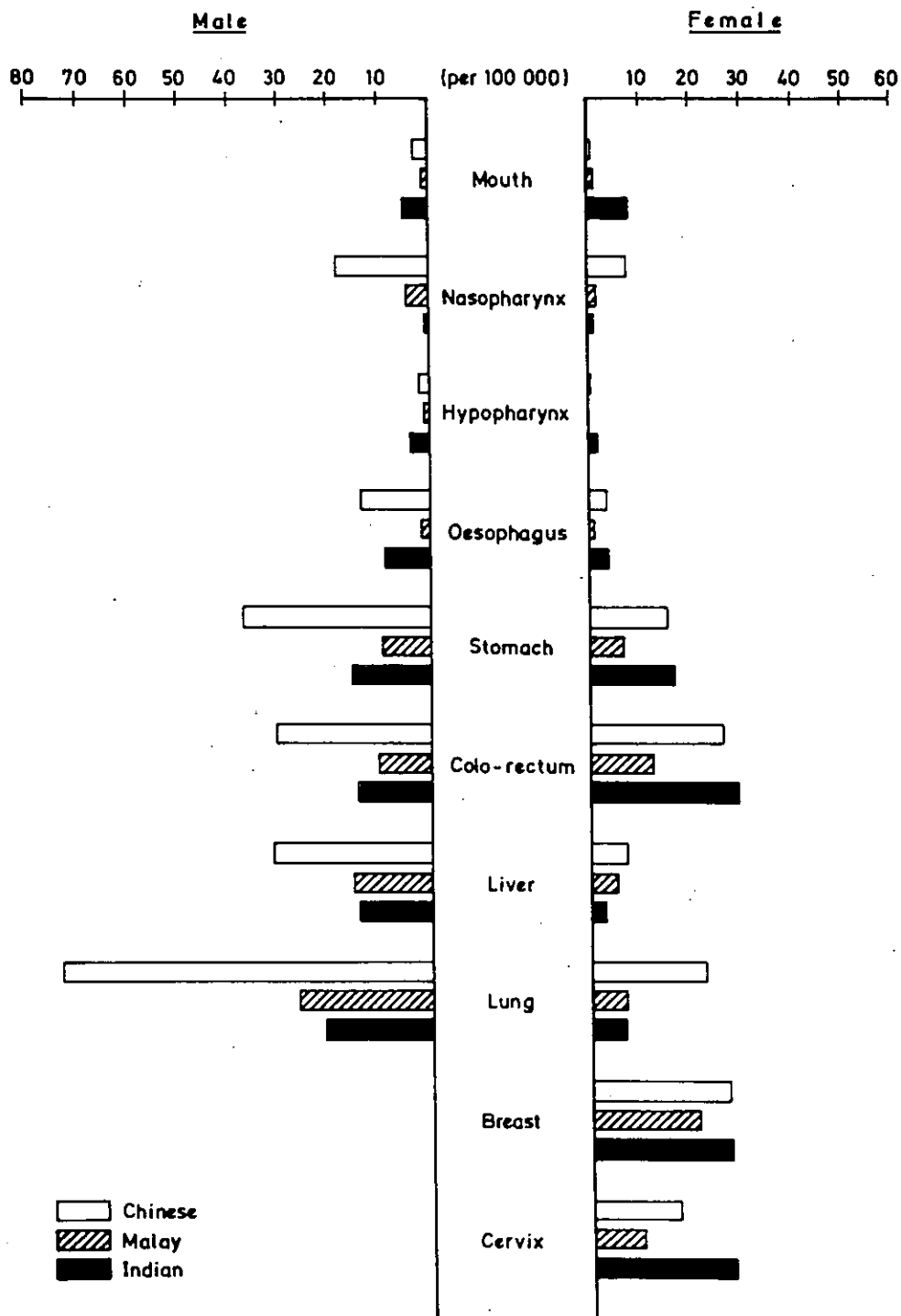


図3. シンガポールにおける人種別がん罹患率の比較
 -中国人、マレー人、インド人の年齢調整罹患率、1978-82年
 (データは Leeら1988年による)。

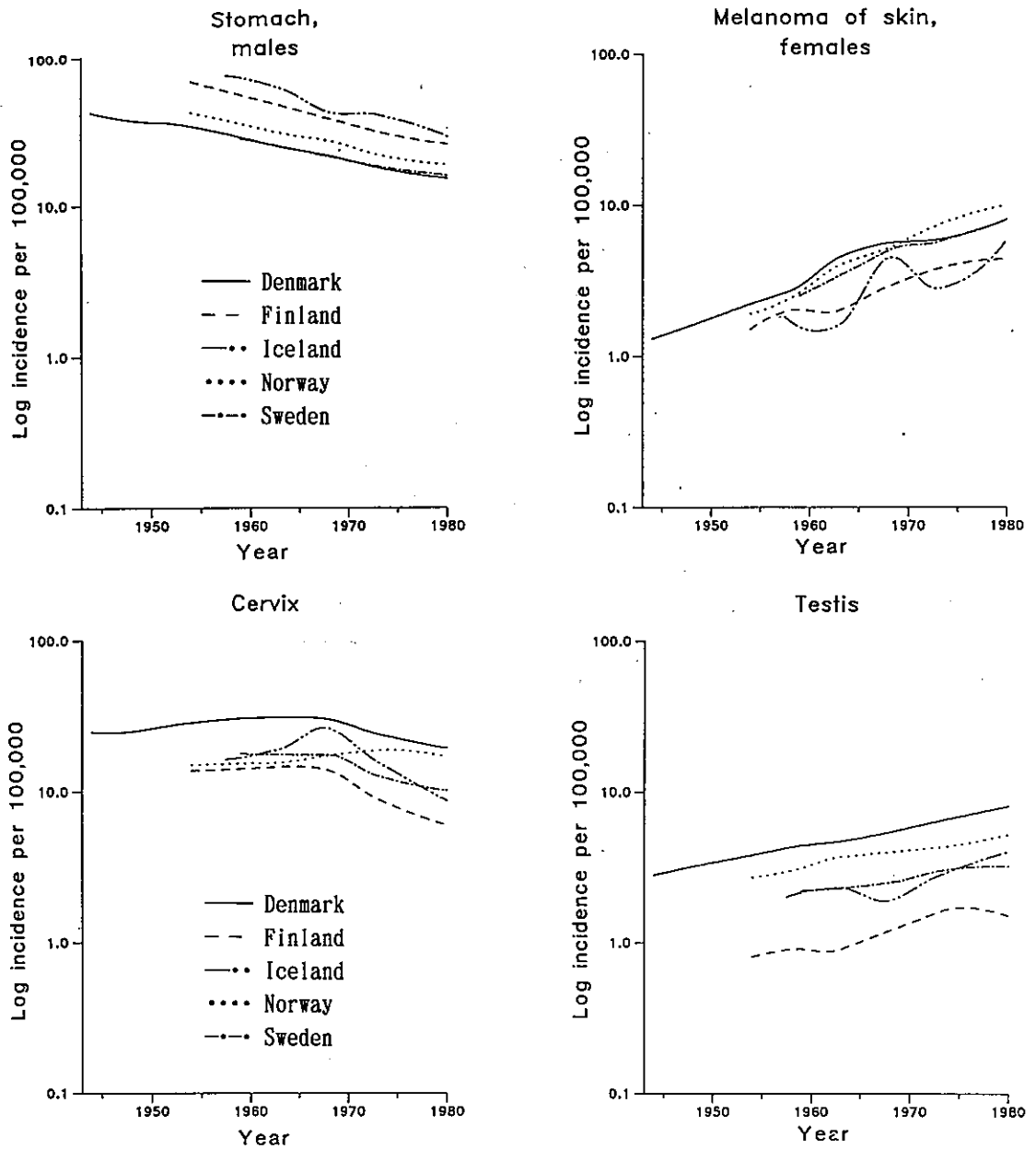


図4. 北欧5ヵ国における4部位のがんの罹患率の年次推移
 一胃、皮膚の黒色腫、子宮頸、睾丸一 (Hakulinen ら, 1986年による)。

(Kjaer et al., 1988) のように、研究努力が報われる世界の地域を指し示すことも可能である。

がん症例は、診断時の居住地により分類することが出来る。これにより、登録地域内での地理的相違を記述することに役立つ。罹患率を、例えば、国または地方自治体別に表にまとめたり、図2に示すように、がんマップに図示することも出来る (Glattre et al., 1985; Kemp et al., 1985; Carstensen & Jensen, 1986; Jensen et al., 1988)。国内の地域が、人口密度によってまとめられる場合もある。その後、罹患率を、例えば、都市部と郡部について (表3)、あるいは、生活様式のような、その他の共通の特性を持った地域について、表にまとめることもできる (Teppo et al., 1980)。

同じ地域に住む異なる民族が、例えば、図3で示すシンガポールのように (Lee et al., 1988)、罹患率の差を示すことがある。世界各地からイスラエルに移民した人々は、がん罹患の大きな違いを示している (Steinitz et al., 1989)。表4で示すように、日本の日本人と米国の日系移民との間でがんのパターンを対比することは、今ではがん発生に関する通常統計から、直接手に入れることが出来る (Muir et al., 1987)。さらに、異なる職業グループ間、社会経済階層間、宗教グループ、あるいは宗教グループ間のそれぞれ、あるいは、組み合わせで、がん罹患率を比較することが出来る。

がん罹患の年次推移を記述し、監視することは、がん登録の大切な課題である (Hakulinen et al., 1986)。なぜなら、例えば、小児白血病、精巣 (睾丸) がん、ホジキン病のように、死亡率は、患者の生存率によって影響を受け、がんのリスクの年次変化を反映しないことがあるからである。年次変化は、多くの地域における皮膚の悪性黒色腫の増加のように

(Jensen & Bolander, 1981)、あるいは、胃がん罹患の減少のように (Jensen, 1982)、その地域におけるリスク要因の影響の変化を指し示すことがある。北欧諸国における年次変化の具体例を図4に示す。がんの推移のモニタリングは、ヘルスケア体系の眼目を計画する

(以下を参照) 場合と同じく、一次及び二次予防対策の評価のためにも、等しく大切である。一次予防の効果 (喫煙のようなリスク要因への暴露の減少) は、罹患の趨勢を観察することにより、最もうまく評価される。一方、二次予防 (例、乳がん検診) の効果を評価する最もよい手段は、死亡統計である。

特定集団でのがん発生に関する統計の生産は、登録それ自体の記録、あるいは、別の目的で収集された他のソースに由来する記録とのリンケイジができる登録では、ずっと促進される。がん登録は腫瘍を記録する。従って登録は、一人における多重がんの発生に関する情報を有している。ある特定の一人における腫瘍記録をリンケイジすることにより、登録は、個人々人における異なったがんの関連性を記述することに多大な役割を演じてきた (Curtis et al., 1985; Storm et al., 1985; Teppo et al., 1985)。がん登録記録と、国勢調査データのような外部データとのリンケイジは、特に北欧諸国 (デンマーク、フィンランド、アイ

スランド、ノルウェー、スウェーデン)のみならず、北米でも、行われてきた。主要な目的は、職業がんの調査であった (Lyng & Thygesen, 1988)。そのようにリンクされたデータファイルは、分子と分母が異なったソースに由来する職業統計よりも明らかに優れている。

分析疫学研究

記述疫学研究に由来する統計成績が因果関係を意味することはあまりない。観察から出現した仮説は、人間における詳細な研究で検討されなければならないし、動物実験で補足されることもある。がん登録は、特定の個人における情報を利用する事が出来るので、分析疫学のための貴重なデータベースとなる。

がん登録記録と他のデータファイルとのリンクの能力は、分析疫学研究におけるがん登録の必須の役割である。このことは、当然、登録と外部データソースとの双方の統一的個人同定指標を必要とする (Acheson, 1967)。がん登録情報は、職業暴露、薬剤服用、喫煙、食餌に関連したリスクを評価するための多くのコホート調査において、エンドポイントとして役立ってきた。登録が長く維持されていればいるほど、カバーする地域が大きければ大きいほど (出来れば一国全体)、そのデータはコホート調査に有益である。

前向き追跡調査におけるがん登録データの利用の場合と同じく、がん登録は、介入研究の結果の評価を促進する。例えば、フィンランドでは、ヘビースモーカーの男性へのベータカロチンとトコフェロールの投与後の、肺、その他の部位のがん罹患が、モニターされた。

がん患者とがんのない対照との間で暴露が比較される症例対照研究は、リスク要因の調査として広く用いられる手法となった。一般論として、がん登録は症例対照研究の実施に適しているとは見なされていない。なぜなら、症例の報告と処理の遅れが、症例の収集が持続的に実施される症例対照研究におけるがん登録の有用性を、限定するからである。がん登録の症例対照研究における主な価値は、そのケースシリーズの完全性と代表性を評価できることである。がん登録は、しかしながら、症例対照研究の貴重な出発点となることを示した。がん登録によって日常収集されるデータを、症例対照研究の手法を用いて分析することができる。これは、とりわけ分母が得られない場合に有益である。例えば、出生地 (Kaldor et al., 1990)、あるいは職業 (Jensen, 1985) に関する情報を使う。

がん登録は、既存の記録、特に病院記録から、暴露に関する情報を引き出すこともできる。なぜなら、がん登録は、しばしば病院病歴ナンバーを記録しているからである。がん登録は、かくして放射線治療と他のがん治療に関連したがんのリスクに関する研究に大きく貢献した (Day & Boice, 1983; Kaldor et al., 1987)。子宮頸がんの患者コホートの中に、症例対照研究を組み込むことにより、そして病歴へのアクセスを獲得する目的で、がん登録の中の患者情報を使うことにより、白血病について、また、多数の固形がんについて、放射線の量反応関係を決定することが可能となった。他の記録から、患者から、もしくは親戚から、暴

露情報を求める調査研究のケース（もしくはコントロール）のソースとして、がん登録を使うこともできる。デンマークがん登録では、フォルムアルデヒド暴露に関連したリスクを調べる目的で、鼻腔がんと他部位のがんとの間で、職業歴が比較された(Olsen & Asnae, 1986)。

ヘルスケア計画とモニタリング

がん登録は、その集団地域における患者数に関する統計情報を提供する。この情報は、がん治療、また、種々のタイプのがんに向けられるケアのための施設の立案と設立に、活用することが出来る。がん発生の地理的な相違が、また、がん罹患の年次推移が、考慮されるかも知れない。図5で示すように、趨勢の変化に関する知識は、さらに、将来の罹患率、患者数の負荷、治療施設の需要の予測のために、用いられる (Hakulinen & Pukkala, 1981)。

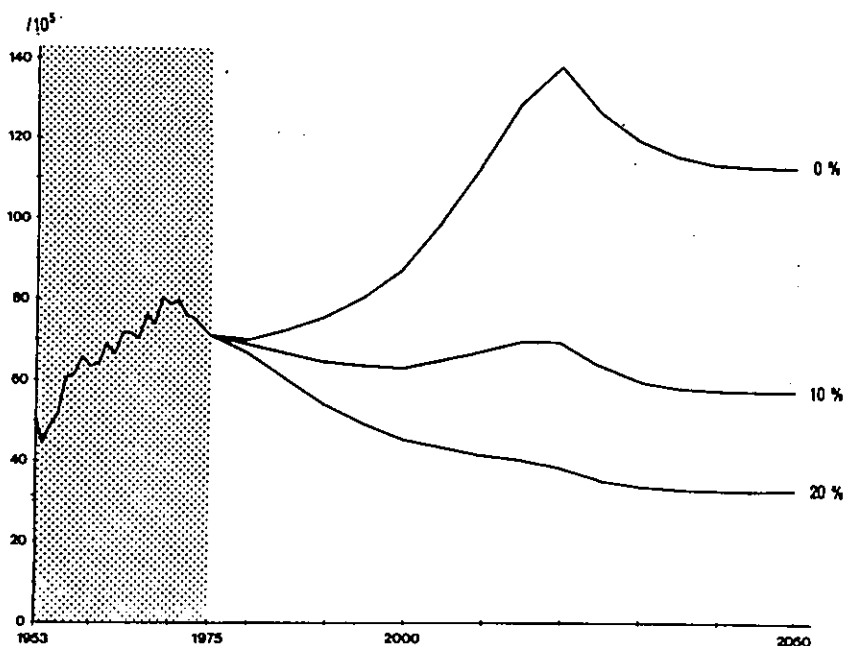


図5. フィンランドにおける1953-75年の男の肺がん罹患率の年次推移と、1980-2050年間の肺がん罹患率の予測（3種）。

将来予測方法：1976-2050年の間、非喫煙者が喫煙を開始する割合を、10-14才では30%、15-19才では15%、20-24才では5%と仮定したうえで、各カテゴリーで禁煙者率を0%、10%、20%と仮定して、3群を作り、5才階級ごとに計算した。

喫煙量は1975年の喫煙者での年齢別成績を用いた。(Hakulinenら, 1986年の報告による)

がん罹患情報は、英国 (Wrighton, 1985) 及びオランダ (Crommelin et al., 1987) において、放射線治療サービスの立案に活用されてきた。イングランド・ウェールズにおける小児がんの罹患と分布に関する知識は、専門化された小児がん関連サービスの立案に有益であることが示された (Wrighton, 1985)。治療施設に関する患者需要の評価は、がん登録のデータと将来推計から導き出すことができる。しかし、より管理的な性質の統計 (例、病床占拠) は、正規には、地域がん登録の視野の外に位置づけられる。より詳しくは、Parkin et al. (1985b) を参照のこと。

患者ケア

個々の患者になされるケアは、ヘルスケアシステムの要素の1つである。がん登録は、患者のケアには間接的にのみ寄与する。例えば、紹介経路を記載したり、あるいは、診断記念日を治療に当たっている医師に通知することにより、患者のフォローを支援することができる。より直接的な寄与は、がん患者に対するケアプログラムの管理であり、それは、ある地域 (例えばスウェーデン) で確立されたが、ある一定のがんの患者が、そのときの最新の診断と治療を受けているかどうかを確認するのに使われている (Moller, 1985)。そうした活動は、特定の新生物を持った患者について、紹介、診断、分類と進行度分類、治療、フォローアップの各方法が、公認されたものかどうかの確認から成り立っている。患者の生存状況のモニタリングは、ケアプログラムに欠くことのできない一要素である。

生存率

殆どのがん登録は、各患者を死亡までフォローアップし、死亡日と死因に関する情報を収集している。患者ケアとヘルスケア計画への重要な間接的貢献は、地域における生存率のモニタリングである (Cancer Registry of Norway, 1980; Hakulinen et al., 1981; Young et al., 1984)。これは、しばしば特殊な病院から得られる詳細な生存率に関する情報を補完することになる。がん登録情報は、図6で示したように、年次動向と同じく、カバーしている地域の、ある特定のサブグループにおける生存率をモニタリングするために使われることもある (例、区別、年齢階級別、性別、社会経済階級別)。もしも真の違いが見つければ、診断及び治療施設が、生存率の低いサブグループに向けられることになる。

がんの治療と生存に対する種々の治療手段の影響は、無作為臨床試験によって最もよく評価される。これらには特別なデザインが要求され、その際のがん登録の役割は、しばしば、新発症例数や進行度分布、地域における生存率に関する情報提供に限られる。がん登録は、そうした臨床試験においてデータ管理や患者のフォローアップを支援することにより、より積極的な役割を演じることもある。こうした情報は、地域がん登録の日常運営として、がん登録に通報されてくるものである。

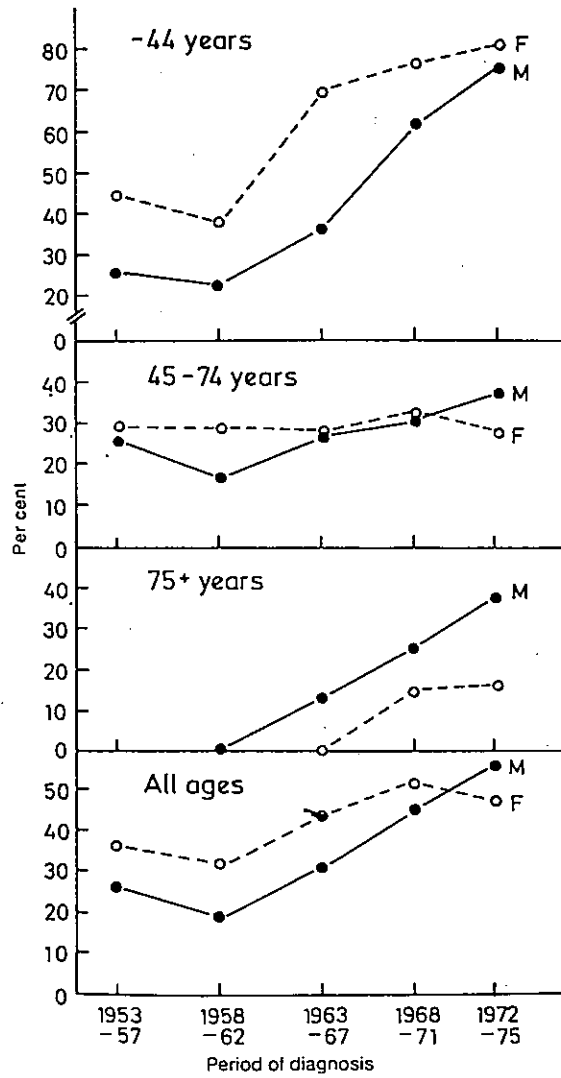


図6. ノルウェイにおけるホジキン病の5年相対生存率の年齢階級別の年次推移 (ノルウェイがん登録室, 1980年による)。

スクリーニング

早期の段階でがんを発見するために無症状の人々を検査することは、ある種の悪性疾患の制圧において、ますます重要になってきている。地域がん登録は、子宮頸がん検診が、様々な普及率で実施されてきた地域間（例、フィンランド対ノルウェー、図4参照）で、頸がんの年次動向を比較することにより、頸がん検診が頸がんの罹患率を低下させる上で有効であることを証明するのに、きわめて重要な役割を果たした（Hakama, 1982; Lynge, 1983; Parkin et al., 1985a）。子宮頸がん検診の効果は、浸潤がん罹患率の年次動向によってモニターすることができるが、他の臓器のがん性病変の早期発見・治療の効果は、例えば、乳がん検診のように、死亡率の動向をモニタリングすることによって評価されなければならない。がん検診の初期のうちには、がん登録は進行度の分布の変化をモニターすることに役立つ。

がん登録のその他の側面

地域がん登録によって収集されたデータの利用の多くは、がん登録業務の欠くことのできない要素の1つである。症例を登録し、登録データを利用することに加えて、地域がん登録は、特殊統計の目的で、がん患者リストを病院部門や研究施設に提供することで、重要なデータソースとなっている。そのような外部研究者による登録データの利用は促進されるべきである。なぜなら、地域がん登録室が、データのすべての局面を探求することは、普通は不可能であり、そしてデータを広めることは、がん登録についての知識とその有用性を増大することになるからである。がん登録データの、強力かつ包括的な利用は、データの質を維持し、改善することにもつながる。

がん登録のスタッフは、しばしば疾病登録や疫学、公衆衛生上の疑問について、相当の専門性を有している。それ故、がん登録は、単にがん疫学だけでなく、疫学的方法論の教育面で、役割を担うことができる。教育目的のためには、地域がん登録は大学及び卒後教育のための材料をもつという利点を有している。

地域がん登録のスタッフは、疾病登録やがんの要因、がん予防とがん患者のケアのプランニングに関する疑問について、保健分野の当局や一般大衆から、助言を求められることもある。

（翻訳：津熊秀明、服部昌和）

第4章 がん登録設置計画

O. M. Jensen and S. Whelan

がん登録の設置理由

世界の人口は約45億人に達しつつあり、その約3/4が発展途上国で生活している。がんによる死亡者数は、世界全体で毎年約400万人と推定されており (Muir & Nectoux, 1982)、1980年には推定635万人の新たながん症例が発生した。これは、人口10万人当り年間143人の罹患率に該当する (Parkin et al., 1988a)。

国レベルであれ地域社会レベルであれ、がん登録は、がんの性質や原因を研究し、保健サービス資源やがん制圧プログラムを計画し、それらの効果を評価するための中心となる (第3章参照)。従って、がん登録活動は、近代的な保健情報システムの一部である。

地域社会におけるがんの実態を把握するために利用できるその他の情報源としては、病院の院内登録データ (第13章参照) や、例えば病理や解剖に基づく特定の患者群の記録がある。そのようなデータから、がんのパターンに関する重要な情報が報告されてきたが、これらはどうしても不完全であり、患者集団の選択的な、偏った標本を表している可能性がある (Parkin, 1986)。がん問題の大きさを正確に描くためには、地域がん登録の設立が必要である。

発展途上国におけるがん登録の役割を、過小評価してはならない。多くの発展途上国では、若い人口が非常に多く、人口の40%以上は15歳未満であり、65歳以上は5%に満たない。がんは、過去においては死亡や障害の原因として重要視されていなかった。しかし、老年人口の増加と、感染症の重要性の相対的な減少に伴い、この状況は変わりつつある。がん登録活動は、がん対策を計画するための情報を提供する、効果的で比較的経済的な方法である。発展途上国におけるがん登録が直面する固有の問題点については、第14章で述べる。

定義

がん登録活動は、地域社会における悪性新生物の影響を評価し、コントロールするのに役立つことを目的として、届出対象の新生物の発生と特徴に関するデータを、継続的かつ体系的に収集する過程と定義できるだろう。がん登録室は、がんの情報を収拾し、蓄積し、解析し、解釈する事務局もしくは機関である。「腫瘍登録」という同義語が、特にアメリカ合衆国で、しばしば用いられている。多くのがん登録では、良性腫瘍や良性疾患、例えば尿路系の乳頭腫や脳腫瘍などをも登録しているので、「腫瘍登録」の方が適切かもしれない (第7章参照)。がん登録資料 (cancer register) という言葉は、がん登録室が保管している情報ファイルや索引を意味する。

症例を登録する方法は大部分同じであるが、地域がん登録と院内登録との違いは明確にし

なければならない。地域がん登録は、特定の人口集団（たいていは地域の全住民）に発生した全ての症例を登録し、疫学や公衆衛生に重点を置いている。院内登録は、その病院における全例を登録し、たいていは基礎となる人口集団についての認識がなく、治療と病院の運営とに重点を置いている。院内登録は、地域がん登録体系の核となりうるであろう。

地域がん登録室の企画

がん登録室を設置する前に、がん登録の目的を明確に定義する必要がある。個々のがん登録室における優先事項は、すでにある医療設備や地域固有の必要性を考慮して、決定されなければならない。地域がん登録室は、特定の人口集団に発生した全がん症例を、長年にわたって収集しなければならない。即ち、がん登録室は、境界を定めた地域の中で運営し、地域住民と外部からきた人々とを区別し、地域外で治療を受けた地域住民の症例をも登録し、重複を避ける為の十分な情報を持ち、その地域にある他の情報源にも十分に接触をもつ。

がん登録室の運営方法は、必然的に、地域の状態と利用できる情報源とによって決まる。がん登録室が発展するためには、広く利用できる医療（制度）と、いつでも診療を受けられる医療施設があって、大多数のがん患者が病気のどこかの時点で保健医療システムと接触を持つようになっていることが必要な条件となる。臨床や、病理のデータを届出する体系が必要であるとともに、信頼性の高い人口集団のデータも利用できる必要がある。地域登録がうまく作動するためには、医学関係者の協力も必須である。必要な費用、職員、設備は時間とともに増大するので、あらかじめ余裕のある企画をたてる必要がある。

顧問委員会

医学関係者の協力と援助を求めることは、設立当初においても重要である。がん登録は、症例の届出を医師に依存しており、また、医師が症例を届出しない場合でも、登録室は診療記録から氏名のある個人についての情報を採録しなければならないので、医師の協力は必須である。がん登録の企画については、医療専門家、医療機関、保健行政機関の委員とともに議論されるべきである。スポンサー、がん症例の情報源、がん登録室のデータを利用する可能性のある人々、からなる顧問機関を設置することは、特に役立つ。

そのような委員会の委員は、国によって異なる。がん登録に出資する機関には、保健医療担当部局、対がん協会、医科大学もしくは総合大学、保険会社、がん研究所、などがある。情報源には、医師会や病院管理部門、病理学や臨床腫瘍学などの特別な専門部門、死亡登録所、および政府の国勢調査局等がある。この委員会は、登録室が設置されて後も、情報源に容易に接触できるように、医療や公衆衛生の各分野との密接な関係を確保するために、ひき続き存続させるべきである。

分母となるべき人口

地域がん登録を企画するにあたり、正確で定期的に発行されている人口データの利用が可能であるかを調査しなければならない。登録対象地域と、登録室が調査する可能性があるサブグループでの、性別、5歳階級別人口が必要である。移民の出入りが把握しがたい国では、国勢調査以外の年の推定人口は、不確かである可能性がある。

がん登録室は、人口動態統計で用いられている定義に従って、同様に、人口グループや地区の定義を使用しなければならない。

法的な側面と情報保護

がん登録室へのがん症例の届出は、篤志で行われるか、あるいは、法律や政令に基づく義務となる。がん登録活動の法的な面は、登録室を企画するときに考慮されなければならない。たいていの国では、がん登録の法律的基盤を保証し、個人のプライバシーの保護を考慮する必要がある。情報保護を考慮に入れることは最重要である。詳細は第15章で述べる。

人口の大きさと症例数

がん登録が対象とするのに最適な人口規模を勧告することは難しい。しかし、実際には、たいていのがん登録室は100~500万の対象人口で運営されている。人口が多くなるに従って、データの完全性や質を維持するのが難しくなるだろう。小さな人口では、意味のある数を得るために時間がかかる。しかし、実際に種々の規模のがん登録室がある。例えば、旧ドイツ民主共和国（東独）は1,700万人、アイスランドは20万人の住民である。

人口の多い国では、イングランドのように、幾つかの独立したがん登録室を設置し、リンクするのが効率的であろう。デンマークのような小さな国では、質の高い国レベルでのがん登録室が可能である。但し、デンマークでは、個人がID番号を持っているので、人口動態統計とよくリンクされていることによる利点もある。全国を対象とすることが難しい国では、合衆国（SEERプログラム）やインド医学研究会議がん登録室ネットワークのように、代表地域に小規模のがん登録室を設置する方がよいかもしれない。

登録室の設置場所

がん登録室を設置する場所は、地域の要因による。がん登録室は、大学、関連病院、保健統計局、あるいは病理学研究所などのように、様々な場所に設置されてきた。

がん登録室が設置される場所は、がん登録室の行政面への依存性に、密接に関係する。効果的に運営するために、がん登録室は、地域の医療施設にある詳細な生死に関する動態情報、医学情報を請求し、かつ、入手することが出来る場所に設置されなければならない。従って、登録室は、何らかの方法で、政府の保健医療担当部門（もし存在すれば）や、専門グループと連携していることが望ましい。対がん協会のような、篤志団体によって設置され、管理さ

れているがん登録室もある。行政的な基盤が何であれ、経験的に、がん登録室は、可能な限り自立しているのがよい。というのは、このことによって、常に成長する組織としてのその時々的重要性を最もよく達成することができるし、他の保健機関との協力を促進し、国内や国際レベルで、直接、関係を持つことができるからである。

財政

登録地域の大きさ、収集するデータ項目数、情報源の数と種類、定期的に登録症例のフォローアップを行うか否か、等が全て、必要な費用に影響する。合衆国のSEERプログラムでは、訓練を受けた登録室の職員が、病院記録から情報を採録する積極的登録を行うとともに、症例を毎年フォローアップする両システムを持っているが、症例一例につき100ドル必要であると評価されている。反対に、アフリカの小さな登録室では、500～2,000人のがん患者を探し、各々の症例について少数の項目を登録するために、1～2人の職員を雇用しているが、年間2～3千ドルで運営されている。フランスのドー登録室では477,671人の人口に年間1,528人の新患者が発生するが、がん登録活動に必要な費用は、3人の肺がん患者の治療費にほぼ等しい。

確かな事実のひとつは、登録活動に必要な費用は、時間とともに増大するということである。年間に新たに登録される患者数が安定していても、積極的なフォローアップをしている登録室では、追跡する症例数は、どんどん増えてくる。データベースが大きくなり、結果の解析や出版を行うようになるにつれて、職員、設備、および場所という意味での資源の追加も必要となる。一端、がん登録室が設置されたら、その後の経済的基盤として、できれば特別研究費を求めることがよいかもしれない。

職員

どんな登録室においても、ただ1つの最も重要な要素は、登録事業が成功するために専念する所長のリーダーシップである。所長は、その他の職員を援助することを求められる。

(1) 数

データを収集し、情報をコード化し、照合し、同じ症例が重なって登録されていないか、必要な情報がそろっているか、矛盾がないか、などをチェックし、解析し、結果を提示するために、職員が必要である。

登録室の開設時から、十分な職員が確保されなければならない。がん登録処理で経験が軽視されがちな面があるので、ここでとりあげた。がん登録活動の方法論は、がん登録室でしか修得できないので、がん登録室の発展につれて、必要数を満たす職員を訓練し、用意する計画が必要である。

費用の面から考えると、必要な職員のレベルと量とは、対象とする人口数、毎年新たに診断される患者数、収集する項目の選択、症例を発見するのに用いる方法、あるいは情報を登

録・コード化・管理する方法によって、大部分きまる（第5－9章参照）。例えば、サマリーを記入した届出票に、場合によっては診療記録、病理や他の報告のコピーなどが添付されているような、自発的な届出に依存する登録室もあれば、症例を選び出し、情報を採録するために、登録室の職員を使う登録室もある。

非常に異なった方法で運営されている4登録室について、付録3a－3cに記載している。テムズ登録室は、大病院には週に数回、小病院にはそのがん患者数に応じた頻度で訪問する巡回職員を雇用している。コロニアのカリ登録室は、主要病院や病理検査部門、X線・血液学・放射線治療部門等から、主に秘書や事務職員を通じて、定期的に症例報告を回収している。しかし、年に一回、必要な訓練を受けた医学生グループによって、全情報源（開業医をも含む）に対する診療録調査も行っている。米国のニューヨーク州では、原則的には、大規模病院の院内がん登録室に届出を依存している。登録室に届出される記録の質を改善するために、院内がん登録に従事する者を対象に、毎年3日間のワークショップを行っている。

各々のシステムによって、必要な登録室職員の種類と数とが異なるので、一般化するのは難しいかも知れない。それでもなお、「5大陸のがん罹患第4巻」にデータを提出した61登録室の実態調査によると、登録室が対象とする人口集団で年に発生する新規患者1,000人につき1人の職員が必要であることが判明した。

（2）質

がん登録室の職員は、専門的、技術的な訓練と経験とをもった職員から構成される。登録室長は疫学と公衆衛生学に関する学識経験と興味とを持ち、腫瘍学についての知識もある医療従事者であるべきと考えられている。登録室の大きさに応じて、病理学、臨床腫瘍学、疫学、公衆衛生学、データ処理法、および統計学のコンサルタントを職員とするか、もしくは助言を受けることが出来るようにする必要がある。

技術職員は、登録職員、即ち症例を発見し情報を採録するのに責任をもつ者や、統計職員、即ち患者の情報をコード化し、腫瘍の記録を処理、加工する者から構成される。がん登録室で必要とされる専門的知識は、実際に従事しながら、もしくは特別な訓練コースで修得できる。データ管理者とプログラマーは、データの保管や利用方法を計画し、実施するために、がん登録室の開始当初から参加する必要がある。

最後に、タイピストや事務官のような職員も、がん登録室の規模に応じて必要となる。

（3）訓練

がん登録職員をあらゆる段階で訓練することは、がん登録室を運営していく上で重要な側面である。がん登録室における作業は、反復的であると同時に、多大な集中力が要求される。全ての種類の職員にとって、たいていは仕事につきながら、特別な訓練が必要とされる。公式的には、一人の職員が自分独自の方法を作ってしまうないように、継続した訓練コースが、全ての登録室に推奨される。同様に、登録室において同じ種類の仕事、例えば、採録しコー

ド化する作業、を行っている職員同士が、十分に議論する時間があることが重要である。自主届出方式を取るのであれば、病院職員のトレーニングコースも企画されるべきである。がん登録従事者へのマニュアルも必須である。これには、米国のSEERプログラムが発行している教本 (Shambaugh et al., 1980a, b, 1985, 1986a, b; Shambaugh & Weiss 1986)、IARC/IACR が特に発展途上国におけるがん登録実務者のために発行した単行本 (Esteban et al., 1991)、あるいはカナダのがん登録実務者のための教本 (Miller, 1988) がある。

がん登録室の職員の訓練を、継続的な基盤に基づいて行うことにより、仕事の精度が上がるだけでなく、職員の確保が容易になる。経験を積んだ職員が退職すると大きな痛手となる。職員に対して、がん登録の目標と目的とを説明し、がん登録室を運営するために職員がどれだけ重要な役割を果たしているかを強調するために、できることは何でもすることが重要である。

設備と広さ

がん登録室を企画するうえで必要な他の側面と同様に、必要な設備と広さも、がん登録室の規模と機能によってきまる。がん登録室は、初めのうちは狭い場所で、少ない設備で開設されがちであるが、将来を見越して余裕を持たせるほうが賢明である。

事務室として通常必要な設備を別にすると、まず必要なのが、症例報告書を保管するための場所と鍵のかかる保管設備である。場所を節約する手段としてマイクロフィルムを用いても、がん登録室が存在しているかぎり、保管場所はいずれ問題となる。手作業で運営している登録室では、ファイルキャビネットのためかなりの場所が必要となる。

登録室が選択するコンピュータ設備も、登録室の規模とその地域の状況による。小さな登録室では、マイクロコンピュータでの運営をはじめている所が多い。病院や大学にある、利用可能な地域の設備を使っている登録室もある。手作業またはコンピュータによる登録室の運営については第8章で述べる。また、発展途上国のためのマイクロコンピュータを利用したシステムについては、本書の付録4で述べられている。がん登録室で利用されているコンピュータ設備の範囲についての情報は、「がん登録室で用いられているコンピュータシステム集」(Menck & Parkin, 1986)で述べられている。

結び

がん登録室を開始することを仮定する際には、統合的ながん制圧プログラムにおけるがん登録室の重要性を強調するべきである。がん登録室の成功は、医療従事者の協力度度によってきまる。地域の医療組織との連携を作り、維持するためには、時間と努力を費やす価値がある。また、がん登録室は長期にわたって運営されるものであることも、同時に心に留めておかなければならない。運営を開始して有効な結果が得られるまでには、数年間かかるかもしれ

れない。がん登録室はその性質から、時間と共に拡大し、必要となる物資上の援助も増大する。従って、登録症例数の増加と、データの利用可能性の増加との、両者の結果としての事業拡大に対応できる、行政的、財政的な将来計画が保証されることが極めて重要である。

(翻訳：味木和喜子、花井 彩)

第5章 データ源と届出票作成

J. Powell

初期評価

データを収集する方法にはいろいろあるが、これらを設定したり、再検討する際には、次の諸点を評価することが極めて重要である。

- (a) 各データ収集方法の実際の費用
- (b) それによって得られるデータの質
- (c) データの利用方法
 - (i) 登録が完了した時点での利用
 - (ii) 長期的（20年以上先）な利用
- (d) 将来の研究に起こりうる制限
 - (i) 項目が全く収集できなかった場合
 - (ii) 項目を省略形で収集した場合
- (e) 情報を提供することにより、各協力団体に生じ得る問題

さらに、これらの要因については、定期的に再評価する必要がある。というのは、新しい方法や必要性が生じて、従来のもものが時代遅れになることがあるからである。

データ源

主要な情報源は、通常、病院やがん診療施設であるが、地域の状況に応じて、地域がん登録は、診療所、開業医、検査室、監察医、ホスピス、健康保険システム、集団健診プログラム、中央住民登録所をも、対象に含める。これらの情報源を全て用いることにより、把握もれの症例がほとんどなくなるのみならず、患者に関連する全ての項目が、1つのファイルにまとめられ、データの質が強化されることになる。しかし、複数の情報源を用いることは、同一のがん症例について、複数の届出が行なわれることを意味する。従って、同一人物に関するデータを効率的に結び付ける方法が、非常に重要となる（第8章参照）。

地域がん登録の作業は、情報を提供する院内登録の協力が得られれば、ずっと簡単になることが明らかである。しかし、このような病院登録がある場合でも、地域がん登録は、さらに他の情報源を利用する必要がある。これは、第1に、登録漏れ（病院を受診しない患者など）を防ぐためであり、第2に、重複登録（例えば、患者が複数の病院を受診する場合など）を同定しやすくするためである。

登録室で利用可能な情報源を以下に述べる。これらの情報源を全て利用することができれば理想的であるが、実際には、それは難しいかも知れない。それでもやはり、最終目標は、

できるだけ多数の情報源を用いることにある。各々のデータ源から情報をえるための実際の方法については、以下のデータ収集方法の節に述べる。

病歴管理室（病歴室）

病歴室ががん登録に寄与しうる方法は多くあるが、それらについては、第13章で述べる。ここでは、登録の完全性と正確性とに影響を与えうる点についてのみ考えることにする。どのような方法を使うのであれ、それぞれの病院に対して登録室は、それぞれ担当者を決めておくことが重要である。この担当者は、その担当する病院からの届出を監視し、量や質の変動に注意する必要がある。その結果、職員（事務や医療の）が変わったり、新しいシステムが導入されたことにより、情報の脱落が生じた場合に、それをできるだけ早い時期に検出することができる。届出が常に良好な病院には余り注意を払わず、問題のある病院に焦点をおくようなことはよく行うことであるが、この結果、前者の病院で重要な職員が辞めたことにより、能力が低下し、登録が十分ではなくなったことに、あとで気づくことがある。

あらゆる段階において、どういう場合に症例を見落とす可能性があるかについて、考慮しておく必要がある。例えば、がんの記録の同定のために、よく用いられる方法は、病歴室が、患者の退院記録を選別し、がん診断（仮の、あるいは確定した）のあるものを選び出しておくことである。この方法の有効性は、全ての記録（病歴、抄録）が病歴室に返送されるかどうかと、病歴室の職員の認識力の効率とによって決まることは、明白である。医師、病棟、あるいは各科に保持されている記録は、病歴室には届かないこともある。このようなことは、少数であっても、大きな問題となる。なぜなら、このような症例は、特定の悪性腫瘍、個々の医師が特に関心をもっているもの、合併疾患、あるいは死亡といった、高度に選択されたものである可能性があるからである。

従って、以下のような問題を考慮することが重要である。

- (1) 特定の長期治療計画の対象患者や、頻繁にフォローアップしている患者では、記録（抄録）が提出されていないかもしれない。これは、特に、造血系悪性腫瘍に当てはまるが、臨床試験の対象となった患者では、化学療法計画が拡大継続される中で、登録されにくいこともある。
- (2) 進行の緩やかな腫瘍の患者では、何年も良悪の診断がつかず、診断がつけられたとしても、治療を行わずに、経過観察のみによって定めるかもしれない。その1例は、目の黒色腫である。
- (3) ある患者では、カルテの中に各種の目的の調査票が何枚もある場合があり、その中のただ1つだけが、詳細な情報を含む場合がある。例：糖尿病など。
- (4) 特定の手術に関する記録を、専門クリニックが自ら保持している場合がある。例：喉頭切除クリニック。
- (5) 移植患者でも、カルテが各科で保持されている可能性がある。患者が後に悪性腫瘍に

なっても、カルテは通常の過程を通らず、届出されないことになる。

- (6) 終末期治療のためだけに入院した患者では、患者の死亡をみとどけるのみであるため、特に、入院してすぐに死亡した場合、カルテにはほとんど注意が払われず、悪性新生物の診断の有無を調べずに、格納されてしまう可能性がある。
- (7) 病歴収納場所が狭くなるにつれて、死亡が判明した患者のカルテは、取り出しにくい公文書保存場所に移される可能性がある。さらに悪い場合は、無秩序に保存され、結果として失われたのと同じになる。

外来患者クリニック

理論上、外来患者は上にあげた情報源のどれかによってカバーされる筈であるが、実際はそうでないことがよくある。

外来診療のみの患者は、生検を受けることがめったにない（従って、この方法からは、患者を同定できない）。記録が不十分である可能性があり、外来患者であるから入院抄録もない。さらに、病院からの定期的な届出には、必ずしも外来患者を含まない。

終末期であったため、患者が入院しなかったり、検査されなかったりした場合でも、もし登録室が死亡診断書を利用できるならば、その症例は死亡診断書から抽出されるだろう（後述の死亡診断書の節を参照）。しかし、合成エストロゲン治療を受けた前立腺がん患者は、たとえがんが存在していても、がんで死亡しない可能性がある。さらに、外来患者としてレーザーや放射線治療をうける患者（婦人科がんや、皮膚がん）の数が増加しており、そのような症例では病理組織報告がないかもしれない。

上記のような患者は、がんという観点では数の上で重要であるが、全体の外来患者数の中では非常に割合が小さいので、問題はさらに複雑となる。従って、これらの患者の同定は非常に困難である。しかし、登録が偏り、不完全になることを避けるためには、これらを含めなければならない。

私立診療所および病院

多くの国では、病院ではなく、むしろ私立の看護ホームや診療所で、診断をうけたり、治療を受ける患者がいる。これらの施設に対しては、届出のために、別の方策が必要である。特に病理報告は、もしそれがなければ、みつけられなかったかもしれない症例を同定するのに、しばしば役に立つ。

しかし、それらの施設の臨床医は、地域の病院の職員を兼ねている可能性が高いので、完全な登録の重要性に気づいているであろう。

病理検査室

登録は、可能な限り、その地域の各病理検査室から（悪性の、あるいは悪性の可能性のあ

る診断の) 病理組織診断報告書のコピーを手に入れるべきであり、これらは、直接、登録室に送られるべきである。

まず最初に、登録室が必要とする記録の同定方法について、病理検査室長と、さらにできれば記録を選ぶ作業をするであろう職員と、話しあうべきである。もし、病理検査室で組織コードが付されていれば、登録対象疾患を区別する最も簡単な方法のひとつとなる。さらにコードシステムのないところでは、例えば、がんあるいは悪性疾患と記載された全ての症例と、登録の対象になる全ての前がん診断の全症例に関し、用語のリストをつくり、登録すべきものについて、意見を一致させるべきである。境界悪性や、臨床的に悪性であるが組織学的にはっきりしないようなものを見おとす危険よりは、「悪性の可能性を否定しえず」のような、疑がわしい症例をも受取って、登録室自身で選び出す方がよい。

心に留めておくべき難しい点は、まず第1に、特定のコードで選択している場合、コーディングミスがあると、症例を見落す可能性があることである。この問題は、前述の用語リストを利用し、更に選択基準を記述する、即ち、コードが特定の範囲内にあるか、あるいは、悪性の記載があるかのどちらかであれば、記録を送るという取り決めになれば、克服できる。

第2は、中枢神経系の良性腫瘍を登録する場合(たいていのがん登録室が一般に実行している)、病理医によって、腫瘍と考えるか、嚢胞と考えるか、意見の異なる場合がある。何を届出で登録するかについて、合意と、届出されるべき診断のリストとがあれば、登録は最終的に非登録対象とするための記録を捜し出す時間の浪費を避けることができるだろう。

院内登録がある場合、病理報告のコピーは、院内登録と地域登録との両方に送られるよう手配をするべきである。これは、一見重複しているようにみえるが、2つの理由で極めて重要である。第1は、長期間にわたって両登録の統一性を維持するためであり、第2には、地域がん登録で、生のデータが利用できることを保証するためである。第1点について、院内登録の職員の移動は、患者の登録効率に直接に影響を及ぼすことになる(未登録の病理報告の集積が、すぐにこれを示唆するだろう)。第2点では、生のデータを利用できることによって、調査の実施可能な範囲と正確性が、かなり広がる。この点は、後述の情報源と収集方法の評価の節で、さらに詳しく述べる。

解剖、骨髄、細胞診報告を含む、全ての種類の病理報告を選別する必要がある。診療所や看護ホームでは、それ自身の病理検査室を持たず、私立の病理検査室を利用しているかもしれない。大きな病院では、別に専門的な病理検査室がある可能性があることも覚えておく必要がある。(口腔病理、神経病理など)

逆に、専門医のいる科は、がん登録が対象としている集団以外からの患者を取り扱うかもしれない。その場合、これらを解析から除くために、大きな注意を払わなければならない。ただし、これらを別に登録することにより、特定の専門医のいる科の標準仕事量や結果を評価するために、役に立つかもしれない。

患者の名前は、間違っつづられていたり、不完全であることもある。そのような場合、

記録は、同じ患者に関する他の文書と正しく対応せず、その結果、重複登録が生じる可能性がある。コンピュータ照合技術が利用できる場所では（第8章参照）、手術日や他の項目の一致状況をみれば、重複を防ぐことができるであろう。

剖検部門

剖検記録は、有益な情報源になる。剖検時に初めて発見された腫瘍が罹患率に与える影響に対し、特別の注意を払うべきである。それが評価できるように、そのような症例には、特別のコードを割り当てるべきである（第6章の19項目参照）。剖検時に見つかった腫瘍の数は、剖検率と同様に、検査を熱心に行った度合を、ある程度反映する。

監察医に報告された死亡についての剖検記録は、監察医のみからしか利用できないような国もあるかもしれない。その場合、この情報源を失わないように、登録室はその地域の全ての監察医と接触すべきである。

血液検査室

血液検査室は、白血病やリンパ腫のような造血器系悪性疾患の重要な情報源である。血液検査の報告書は、通常、固形腫瘍とは別の検査室からくるので、血液検査記録のコピーが登録室に届くように、病理検査室とは別の手配をすることが重要である。必要とする用語のリストを、拡大する必要があるかもしれず、境界診断例が多くあると思われる。従って、対象に含める疾患についての打合せと、厳密な定義が必須である。細胞診の報告も含められるべきである。

治療も行う臨床病理医が検査室にいる場合は、登録室が必要とする治療の詳細や他の項目を得るために、その医師の協力を得るべきである。ここでもカルテは検査室に長期間保管されていることがあり、病歴室ではみつからないかもしれない。

他の検査室

がんの診断に役立つ種々の生化学的あるいは免疫学的検査は、上記以外の検査部門で行われるかもしれない。例えば、血清の酸性フォスファターゼ（前立腺がん）、血清アルファフェトプロテイン（肝細胞がん）、血漿蛋白のパターン（多発性骨髄腫）の測定などがある。その他の腫瘍に特異的な抗体で現在用いられているものは、特定のがんに対する診断特異性が低いが、将来には、がん登録室にとって有益な情報源になるかもしれない。

全ての検査部門と同様に、ここでも個人同定データは、必ずしも正確で充分であるとは限らない。同定項目の記載ミスや不十分さ（例えばデータがない）は、結果として中央登録での重複登録を引き起こすかもしれない。

死亡診断書

症例同定のための非常に重要な情報源として、死因欄に悪性疾患の記載のある死亡診断書がある。多くの国では、死亡を登録する体系があるが、情報保護の理由から、診断は取り外しのできる用紙に記載されているところもあり、もし、名前や他の同定情報の部分から切り離されていれば、この情報源は、がん登録には役に立たなくなる。

死亡診断書の最初のモデルは、WHOによって1948年に考案された。多くの国が、このモデルを採用し、直接死因、原死因、および死亡には直接関係しないが死亡時に存在した他の病理的状态、をそれぞれ区分して記載する原則を守りながら、各々の必要に合うように改訂した。

死因は、1948年から、国際的に同意された規則を用い、国際疾病分類（ICD）に従ってコード化されている。結果の出版形式にも細かい規則があり、原死因について集計される。

これらの項目を、がんについての情報源という観点からみると、このシステムの原則的な目的、即ち、死因を集計することには、困難な問題を生ずることもある。というのは、他の疾病や、事故の結果で死亡したがん患者では、死亡診断書にがんが記載されていることもあり、記載されていないこともあるからである。

死亡診断書の原本（あるいは原本のコピー）を吟味することは、人口動態局の診断リストに依存するよりも、はるかに望ましい。後者は原死因についてのみコード化するので、がんが原死因でなかった死亡は、含まれないかもしれない。がん登録が必要とする情報を得るためには、死因についての詳細なデータは、重要である。

死因の診断は、しばしばあいまいな用語で示されている。悪性疾患についていうと、部位は記載されていることが非常に多いが、特に老人の患者の場合には、これが必ずしも正確であるとは限らない。しかし、死亡診断書は、通常、行政当局そのものが作り上げるので、生年月日、死亡日、住所などの個人同定項目については、通常正確である。これらの要素は、もし生存率解析が登録の目的のひとつとして行われるのであれば、特に重要である。

死亡についての情報は、地域がん登録にとって、常に重要なものである。以前に登録されていないがん死亡患者が発見されることは、よくあることであり、遡及調査が開始されなければならない。

(1) 病院で死亡したり、剖検を受けた患者の死亡診断書については、病院あるいは病理医に、患者の臨床抄録を要求する。

(2) 病院以外で死亡した患者については、請求は、死亡を診断した医師に送られるべきである。これについては、次の節で詳細に述べる。登録の様式よりも、電話の方に、よく対応してくれる医師がいるかもしれない。終末期のみ診療した医師から患者は入院したことがないという情報をえたとしても、後になって、不正確であったことがわかるかもしれない。地域がん登録では、たとえ追加情報が全くなくても、全ての症例を登録することが推奨される。しかし、死亡診断書以外の情報のない症例は、「死亡診断書のみで登録されたもの」として同定できることが重要である。

死亡診断書にがんの診断があって登録されたが、後になってその診断が間違っていたことがわかった症例（臨床記録の遡及調査や剖検によって）は、除かれるのが最もよい。もし残しておくのであれば、特別な旗を立てておき、罹患症例の解析からは除くべきである。

死亡診断書のみで登録された患者の割合は、しばしば、登録過程の質の指標として用いられる。がん登録室は、上記の死亡診断書のみのもので定義、即ち、遡及調査がうまくいなくて、腫瘍の証拠が死亡診断書のみであるような症例、に用いるように推奨される。この共通の定義は、登録室間の比較性を高めることになる。しかし、がん登録室は、死亡診断書から初めて登録室に認識された患者の数を、監視すべきである。なぜなら、そのような数が多い、あるいは増加することは、届出システムが不十分であることを示唆するからである（第9章参照）。

死亡診断書のみで認知された症例を、別に解析することが有意義なことがある。これらの症例は、老人、ある民族、あるいはある宗教集団のような、特殊な集団に属する人であるかも知れない。このような場合には、医療サービスを拒否していることもある。そのような集団で観察された罹患率（実際は診断率である）の妥当性には、疑問がある数値となるであろう。

死亡診断書によって、初めて登録室が見つけた症例の遡及調査は、登録室が登録活動を開始した時の有病症例を除くためにも、非常に重要である。しかし、有病症例を偶発的に含むことは、ある程度、避けがたいことである。そのような症例が多い場合は、登録開始年、あるいは2年目のデータは、公表することを避ける必要があるかもしれない。

開業医

開業医は、しばしば、がん患者を診察し、病気が悪性であることを疑う最初の医師である。たいていの先進国では、がんの疑いがあるとすぐに、開業医は患者を病院やがんセンターに送る。開業医が利用できる情報は、病気の最初の症状、患者の既往、家族歴以外は、時折、限定される。

患者は、初診時に既に、何の治療も役に立たない程、がんが進行した状態であるかもしれない。医師は、診断価値の余りない、苦痛を伴うような検査をしないように決定するかもしれない。このことは、特に、先進国での老人や、発展途上国での全ての年齢層にみられる。開業医は、そのような患者についての唯一の情報源であり、地域がん登録室におけるこれらの症例の主要な情報源である死亡診断書を、通常作成・記載する医師となる。

前章で述べたように、未登録患者の死亡診断書を受け取ると、登録室は、その患者について最少限、かつ十分な情報を尋ねて手紙を書くべきである。手紙を書くタイミングは、地域の状況に依る。開業医が持っている記録を中央事務局に戻すまでに、できるだけ早く情報を求めることが重要である国もある。しかし、例えば2か月待つことにより、開業医に接触する必要なしに、病院から必要な情報を受け取ることもあるかもしれない。

別の方法は、全ての開業医に、患者が病院に紹介されなかった場合に使う届出票を受取人払いの封筒とともに、渡しておくことである。届出票には、患者が入院しなかった理由（高齢、拒否など）とともに、症例の登録に必要、かつ十分な情報を含める。将来的には、マイクロコンピュータを持つ開業医が増えるに従い、そのような患者のリストを定期的に作成することが可能となるであろう。

健康保険（労働者補償基金など）

健康保険システムは、全面的な国のサービスとして、又は、人口の相当部分に対する強制保険として、或は任意保険として、多くの国で発展してきた。

そのようなシステムでは、被保険者への給付金の払い戻しに関連する管理的な記録に重きをおいており、医学的情報は、乏しく、あまり正確でないかもしれない。その一方、同定項目に関する情報、即ち、正確な氏名、生年月日、住所、職業歴が得られる。この観点からみると、国によっては、医療情報は満足できなくても、患者情報を確認するために保険機構が重要な情報源となるであろう。健康保険機構は、被保険者に関する全ての記録を収集する仕事をしているので、ある環境下では、健康保険機構は、種々の情報源とがん登録との仲介としての役目を果たす。この場合、情報保護の障害が克服できれば、これらの機構は非常に価値のある情報源となる。

これらの体系には、登録の観点からみると、ひとつの大きな不都合な点がある。それは、同定データは被保険者のものしかないが、患者は扶養者、例えば配偶者であるようなことがあるからである。

スクリーニングプログラム^{*1}

がんをできるだけ早く発見するためのスクリーニングプログラムが、最近30年の間に設立されてきた。主なプログラムは、子宮頸がん和乳がんを目的としていたが、アニリン染料産業の労働者における膀胱がんのように、他のがんを発見し、検査するためにも組織されてきている。そのようなプログラムでは、それによって発見されたがん症例についての詳細な情報を、そのプログラムの組織団体が所有している。これらのプログラムから情報を得ることは、一般に容易であるが、浸潤がんを、上皮内がんや他の前がん病変から区別するためには、通常さらに、どこかで検査する必要がある。スクリーニングプログラムからのデータを含む影響は、例えば生存率を計算する場合などには、注意して評価する必要がある。さらに、症状のない人をスクリーニングすることによって、臨床的には診断する機会のない腫瘍を発見することになるかもしれない。従って、スクリーニングで発見された症例をがん登録に含むことにより、罹患率が見かけの上で増加するかもしれない。

^{*1} 集団健診などの健康診査プログラム（訳者注）

その他の疾患の発見体系も、重要な情報源となるかもしれない。即ち、X線検査による肺結核の検査は、結果として肺や縦隔のがんを発見する。

スクリーニングプログラムは、症例の情報源として価値があるが、罹患率の比較における影響を念頭におく必要がある。もし（発見の方法（第6章の19項目）という情報を用いて）スクリーニングで発見された症例が区別できるのであれば、その様な症例を含んだ、あるいは除いた罹患率を計算することができる。

中央住民登録

多くの国では、全人口を対象とした中央住民登録局がある。全国民総背番号制の国や、各人の詳細内容を持つ中央アルファベット索引ファイルのある国では、この登録が最も重要なものとなる。病院記録よりも正確であるため、そのような行政の住民登録からえられる個人同定や人口学的な情報（性、生年月日など）を登録することは、利益があるだろう。中央住民登録は、ある登録地域から別の地域に移動した患者を追跡するために使うことができる。また、高危険群の人々に旗をたてる（目印をつける）ために使用することもできる。この点に関しては、第3章で述べる。

ホスピス

終末期の疾病に対して、これらのホームが患者のケアにおいて果たす役割は、増加してきている。一般に、患者のほとんどは、病院で診察を受けているが、それでも、ホスピスからの届出で、以前に発見されていなかった症例が同定されることがある。

そのようなホスピスの大多数は、事務的な能力が比較的少ないため、情報の要請は、最小限にするべきである。しかし、死亡診断書が受け取られ、その症例が登録されていなかった場合、ホスピスは、簡単な届出票であれば、通常喜んで記載してくれるだろう。住所と、最初に診断された、あるいは治療された日付と場所とを含めることが、特に重要である。他の情報が得られない時には、これらの項目は、まず第1に、地域住民であれば、症例を登録に含めること、第2に、適度に正確な日付がえられること（例えば有病ではなく罹患症例であることを確信できること）を確実にするのに役立つであろう。

老人の長期滞在病院あるいは老人ホーム

このようなタイプの病院からも、患者の届出を得られるよう依頼しておく必要がある。

データ収集の方法

この章の前節では、登録室が利用できる情報源について述べた。これからは、通常行われている医療記録の一般的な面と、さらにそれぞれの情報源が登録室に情報を伝達する方法に

ついて述べる。伝統的に、届出方法は、能動的と受動的とに分類されてきた。

能動的な届出（情報源施設での情報収集）では、実際にデータ源を訪れ、所定の届出票に必要な情報を記入する、あるいは必要な書類のコピーを得るための、登録室の職員を必要とする。

受動的な（あるいは自主的な）届出では、届出票を記載し登録室に送るか、あるいは、必要な情報のある退院抄録などのコピーを登録室に送るための他のヘルスケア従事者を必要とする。

実際、これら2つの方法を合わせて使用することもできる。例えば、能動的な病院訪問を、病理検査報告のコピーと、がんに関する記載のある死亡診断書とを受動的に受取ることで補うなど。

病院における日常用の医療記録は、基本的な情報源として特に役に立つ。しかし、これは一般にがん登録室の需要に合致しない。というのは、がん患者は、全入院患者の中の少数を占めるにすぎないからである。がん登録室が特に関心をもつ個々の項目を知らなかったり、また、ほとんど注意を払わない可能性がある。とりわけ、病院システムは患者の入院医療に基本を置いているが、地域がん登録は、腫瘍の入院医療に関心をもつ、即ち、腫瘍の経過についての全ての情報をまとめること、特に、同じ患者が繰り返して入院することによる重複登録がないようにする。

地域登録では、病院の入院から退院までの管理方法と過程について、また、既存のファイリングシステムについて、よく理解していることが重要である。入院係事務職員は、同定項目を記録し、その正確性と完全性とに責任をもつ。患者の診療録ファイルの管理に責任を持つ人は、看護婦あるいは病棟秘書、各外来科の看護婦や秘書、あるいは全科に関係する病歴室の病歴管理士である。診療録ファイリングシステムに関しては、何度も入院した患者の病歴は1つのファイルにまとめられているか否か、また、その病院で診断を受けた全てのがん症例についての届出が完全に行われているか否かをチェックするために、そのファイリングシステムをどのように使うことができるか、などがわかっていなければならない。緊急入院の症例では、最小限のデータしか利用できないかもしれない。がん患者が、腸閉塞や悪性胃潰瘍の穿孔により、緊急入院する可能性があることは、よく認識されていない。

これらの詳細な知識に基づいて、登録室は収集する情報項目を決定し（第6章参照）、さらに、地域の必要性に沿った情報を収集し、しかも腫瘍登録室にとって質の高い情報が提供できる体系を、考案することができる。

病歴室は、しばしば、主要な情報源となりうる。登録室が利用することのできる主要な資料源には、次のようなものがある。

特別に作成された登録票

放射線療法記録あるいはその要約のコピー

退院紹介文書や症例要約のコピー

病院患者情報システム

これらについてのデータ収録票は、必ずしも、一つに限る必要はない。例えば、放射線治療科でみた患者については放射線療法記録のコピーを用い、他の全ての患者については届出票を用いる方式も、場合によっては有益である。なぜなら、後者は、がん以外の多くの他の疾患を扱う診療所で診断を受けるだろうから。

将来的に、これらの情報源は、紙ではなく、コンピュータ・メディアで登録されることが増加するであろう。

特別に作成された登録（届出／報告）票

これらは、登録室が作成するが、登録室が必要とする同定内容と臨床内容の詳細についての要約情報を提供する様式である。これらは、病院の職員（主治医、登録士、秘書、病棟事務員、病歴士等）によって、あるいは、登録室からの派遣職員によって記入される。この種の届出の大きな利点は、必要な情報がその様式に限定されているため、標準的なデータを提供できることである。また、資料を比較的速く点検することができるので、記載もれを早く発見し、処置することができることである。

病院職員が記載し届出票を受け取ることが主要な登録方法である場合には、病院では余分な仕事とみなされ、その精度に大きなバラツキがあるという短所がある。一般に、医学的に訓練された職員は、彼らの医学的知識に拘わらず、このような届出票を記載するには最適の者ではないということ認識しなければならない。まず、彼らは十分な注意をもって届出票を記載するだけの十分な時間がなく、またその意欲が少ないかもしれない。第2に、職員は、非常にしばしば転勤するため、一貫性という重要な要素が失われる。後者の点は、医学職員以外にもいえることであり、職員の異動が起こった場合には、新しく配置された職員を、登録室、あるいは病院で、訓練する努力をしなければならない。

登録室が雇用する派遣職員が複数の病院を訪問する長所は、登録室が職員を直接制御でき、返信の質を簡単に監視できることである。これらの職員は、連絡係として、病院レベルで必要とする訓練を実施し、名前の書き間違いのような問題を明らかにしたり、また最も重要なことであるが、病院の職員に（臨床、事務ともに）情報源としての登録の重要性に気づかせる、非常に有益な機能を果たすことができる。こうして情報の還元が奨励される。リストの形であれ、より高度な解析をした形であれ、これがデータを改善するための最も効果的な方法となる。

不運なことに、情報の質という観点での費用効果についての正確なデータはまだないが、派遣職員は、比較的到高価である。また、外部の職員を用いることは、病院職員の参加と関わり合いを少なくするという不利を招く可能性がある。

放射線治療記録あるいは要約のコピー

これらが利用できれば、これは、登録室にデータを提出する最も効率のよい方法となる。放射線治療科は、ほとんどがん患者を対象としているので、情報は、登録室が必要とするものに関連することが多いだろう。よくある問題点としては、以前の検査や外科治療についての情報が不十分であることであるが、そのような情報は、地域登録としては、収集していないものかもしれない。さらに、初発時に外科治療を受け、再発や転移の治療のために紹介された患者が多いかもしれない。有病症例を含むことにより、罹患率が変動することを避けるために、これらを区別することが重要である。

退院紹介文書や症例要約のコピー

財政的問題のため、特別の届出票を記載することは、（これは望ましいことではないが）不可能になることがあるかもしれない。そのような場合には、一連の治療の終了後に送られる退院紹介文書や退院サマリーを調べることも価値があるだろう。その精度が十分なものであり、質のよいものが多ければ、カーボンコピーや写真転写で必要な情報の多くを得ることができる。

抜けている項目は、普通、管理的項目である。通常、退院サマリーには省略されている詳細内容、例えば、個人同定項目の詳細や、職業などの項目のみを含む省略型の届出票を考案することで、これらを得ることができる。この届出票は、登録室に届出する前に、退院サマリーに付けることができる。

この方法の長所は、

- 転記ミスを除ける
- 特定の届出票を使って採録する労働力に比して、必要費用が小さい

短所は、

- 個々の臨床医によって退院サマリーの内容の精度がきまるので、データの量と質が一定しにくい
- 臨床以外の項目は、十分詳細にもらえないかもしれない
- 一般に、多忙な業務の間に、多種類の疾患のサマリーを扱うことになり、多くの秘書や事務職員をわずらわせる
- そのような状況では、悪性症例の選別は、計画性を欠き、見おとす症例もでてくるかもしれない
- コピーをとる実際の仕組みにも、困難な点があるかもしれない。例えば、写真複写機のある場所が遠いとか、カーボンコピーを使う場合、医師が吹き込んだテープをタイプピストがタイプで再生する前に、その診断書が悪性疾患の1つであることを知らず、カーボンコピーをとっていない場合があるかもしれない、などのことがありうる。

病院患者情報システム

病院が、その活動や労働量を測定する方法はたくさんあるが、そのようなシステムの目的は、全ての医療内容を含むことだけでなく、その処理をできるだけ手早くすることにあるので、登録室は、これらのデータとしての可能性を無視してはならない。

患者退院時、あるいは各々の医療の終了時に、情報をコード化することにより、労働量を監視している病院では、これは、悪性症例を早く同定するのに貴重な方法となりうる。患者が選別されれば直ぐに、記載された様式のコピーを取ったり、各々の患者リストを打ち出したり、登録室が使用しているコンピュータへデータを電送することにより、情報を登録室に転送することができる。

前述のように、この種類の日常業務上の医療書類は、洗練されてはいるが、がん登録には不十分であることが多い。しかし、このシステムは、次の2つの方法の1つとして利用できる。

(1) 日常業務上の書類は、正確な個人同定項目、詳細な管理項目（氏名、住所、性別、年齢、病院、病院コード、担当医師名）や、仮診断名、およびおそらく治療方法の指示をも提供し、初期登録に用いることができる。

電算処理され、コード化されたこれらの情報は、元のデータに追加、補足されるべきである。補足する情報の範囲は、登録室が提供することを要求されているサービスのレベルと、間接的に、地域における研究活動に依存する。必要な項目としては、原発部位、進行度、手術の詳細、および補助治療についての記述を含むであろう。この臨床的な情報が、この章で述べたどれかの方法で、病理検査報告のコピーなどとともに、登録室に送られれば、中央でのコード化を保証することとなる。将来の研究にも、データの確認過程においても、登録室で基礎情報が利用できることを保証することも、同様に重要なことである。

(2) このレベルのサービスや研究が必要とされていない、あるいは実現不可能である場合には、日常業務上の情報は、ずっと少ないレベルでの情報で補足することで済む。この方法では、コード化はほとんど登録室外で、日常業務として病院データ抄録作成をする事務員によって行われる。臨床的な追加情報は、届出票や、コンピュータの打ち出しリストで補足されるが、これは抽出されたものであり、それ故、省略されたものである。従って、大がかりな妥当性の検査は不可能である。さらに、速さに重点がおかれているので、診断の修正はあまり行われておらず、退院時では、その患者の全ての検査結果がまだ利用できない可能性もあるので、かなりのエラーの元になりうる。

しかし、資金の点で、詳細な情報を収集することができないのであれば、これは、速く、比較的費用のかからない登録方法である。症例が死亡時に同定されるのみよりは、確かにより正確である。欠点は、登録室外でコード化されていることと、将来の研究のために必要な詳細なデータと原材料が欠除していることにある。後者の点については、原本の病院カルテが消失することなどがあれば、さらに重要な問題となるであろう。

上述の両方法における、最大の障壁は、繰り返し入院である。同じ患者についての繰り返し入院や繰り返し受診の記録を一本化する何らかの方法を利用することが必要である。さらに、この結合は、どれだけ時間が経過しても行うことが必要である。コンピュータ化された所ではコンピュータが導入される前の入院をも考慮に入れることが極めて重要である。そうでなければ、転移で診療をうけた患者は、初期治療から何年もたっている場合が多いが、こうした患者が、新規の症例として登録されるであろう。結果として、それらは、有病率にのみ含まれるべき症例なのに、罹患率をも不当に増加させる。

届出要綱

どの方法を用いるにしても、各施設が、がん症例を届け出る規則と要綱とを認知していることは賢明である。これらは、届出票に印刷したり、特別の手引きとして添付することができる。その詳細は、登録室が必要とするデータによって、大きく異なる。下記の条件は、ドイツ登録室の手引きに基づく。

- (1) 届出対象疾患のリスト。これは、以下の形式となるであろう。
 - (a) 必要とする国際疾病分類 (ICD) の範囲、例えば、140-208
 - (b) 届出する施設が ICD コードを用いていなければ、実際の用語のリスト
- (2) 届出対象例のリスト
 - (a) 新たに腫瘍と診断された全症例
 - (b) 多重がんの全症例、それぞれの腫瘍について1つの届出
 - (c) 診断が、届出対象腫瘍の範囲内で変更された場合
 - (d) 以前に届出された腫瘍が、届出対象外の疾患に診断変更した場合
 - (e) 前がん病変や上皮内がんが、浸潤がんに進展した場合
 - (f) 原発がん診断後4か月以内に治療が変更した場合
- (3) 届出に責任をもつ人
 - (a) その腫瘍が以前に別の科から届出されたか否かによらず、その科が、始めて臨床的あるいは病理組織学的に登録対象疾患を診断、管理あるいは治療した時に、その病院の科の主任は届出する義務をもつ。
 - (b) 病院に紹介せずに、届出対象の腫瘍の治療を開始するか、管理している開業医あるいは専門医
 - (c) 老人などの公共施設やホームの医師、届出対象腫瘍の診断をしたが、病院に紹介しなかった場合
 - (d) 病理講座の主任が、剖検で届出対象の腫瘍を診断した場合、あるいは、以前に届出対象疾患であることが疑われていたが、剖検で見つからなかった場合
- (4) 届出票の記載方法のガイドライン

(5) 何か問題のあった場合に連絡すべき登録室の担当者名

(6) 結果（リストや解析など）について連絡をとるべき登録室の担当者名

情報源と方法の評価

データを収集する究極の目的は、それを利用することであることを忘れてはならない。この理由から、各々の情報源が登録の正確性と完全性とに与える影響についてのみならず、その後の解析や研究における有効性に与える影響に関しても、評価されなければならない。このことは、コンピュータ化が進むについて、特に重要となる。というのは、個人同定指標は別として、コンピュータ入力したデータは、ほとんど必ず、情報をコード化するか、文字を圧縮することによって、単純化することが必要とされているからである。

これらの作業が必然的にエラーの発生の危険をもつことはよく知られていることであり、この点に関しては、第9章で述べる。もし元データが登録室で利用できなければ、これらの作業が、どの程度将来の研究を危うくするかは、あまり明確でなく、しばしば忘れられている。

例えば、登録室では、まず組織型を（例えば、国際疾病分類腫瘍学、ICD-0 によって）コード化するにあたり、病理検査報告を用いるが、病理学者によって記載された報告には、しばしば、深達度やリンパ節への拡がりについても記載されている。初期の解析では、ICD-0 にコード化している情報しか必要としないかも知れない。しかし、その後の研究では、例えば、深達度の予後への影響の評価が必要とされるかもしれない。もし記録のコピーが登録室に保管されていれば、必要ならば長期間にわたるそのような研究も実施できる。しかし、もしコード化が登録室の外で行われ、既にコード化されたデータを記入した届出票や磁気テープに依存しているのであれば、そのような将来の発展的利用は不可能である。

さらに、元データが利用できない場合、あるいは、情報が登録室外でコード化され、磁気テープで提出されて元データの利用が制限される場合は、ある部位を見直す際に、（現代の情報処理で）実現可能な妥当性のチェックができない。情報が、既加工された形（磁気テープやカセット）で登録室に受け取られるならば、検査したり、さらにコード化を上げるといったオプションもない。

がん登録は、今日のみならず、明日のがん制圧と研究のために、データを収集する。従って、データの収集方法を決定する時には、各方法を選択した場合の影響を、細心の注意をはらって考慮することが重要である。間違いや脱落が、その後修正される機会は、殆どない。

（翻訳：味木和喜子、馬淵清彦）

第6章 がん登録において収集する患者情報項目

R. MacLennan

がん登録に必要な情報は、がん登録室の機能と直接関連しており、その機能によって決定される。院内がん登録事業は、根本的にはその病院を訪れたがん患者についての調査であり、これについては第13章で別に取り扱った。一般地域集団におけるがんの罹患を正確に評価できるのは、地域がん登録事業において他にない。必要な情報項目は、それぞれのがん登録事業の機能と目的を十分考慮してからでないとは完全に特定できないが、ほとんどすべてのがん登録事業に共通の基本項目がある。

がん登録の基本情報項目

患者管理に関わる登録（注：院内登録など）では必要不可欠な情報項目でも、がん罹患の推計が第一の目的である地域がん登録では、明らかに不要な項目が多い。「基本情報」という用語は、一般にすべてのがん登録室で収集されている項目を指す。それ以外の項目を収集するかどうかは、その登録の目的、情報の収集方法（第5章参照）、あるいはその登録室が利用可能な情報源によって定まる。登録室が収集する項目と、登録室がデータとして蓄積する項目とを区別することが重要である。収集した項目すべてをコード化して蓄積するわけではない（例えば、管理目的で利用される情報項目が、それに当たる）。

どのがん登録事業でも基本となる情報項目を、表1に示す。あまりにも多くの項目を収集しようとして失敗したがん登録室は多い。情報の量よりむしろ質の方に重点をおかなければならない。もっとも成功し成果をあげた登録室の中には、個々の患者については非常に限られた情報しか収集していなかった例もある。これらの基本情報項目は、どこの地域がん登録室にとっても意味のあるもので、発展途上国で収集するのはこれらの項目のみでよいくらいである（第14章参照）。

任意追加項目

情報項目の追加は、登録の処理を複雑にするとともに、経費を増大させる。したがって、追加する項目一つ一つについて、「その項目を収集したいか」を問うより先に、「なぜその項目が必要か」、あるいは「その項目を収集するための経費があるか」を問わなければならない。追加項目を網羅して表2に示した。消化管や小児腫瘍登録室のような特定研究用の登録では、その専門分野の利用者の責任において定めた追加情報項目群を収集、コーディング

表 1. がん登録の基礎情報

項目番号	項目	注釈
個人		
	個人同定	
3	氏名	その地域の習慣による
4	性別	
5	生年月日または年齢	不明の場合は推定する
	人口統計情報	
6	住所	ふだんの住所
11	民族	地域に複数の民族が存在する場合
腫瘍		
16	罹患年月日	
17	診断の根拠	
20	解剖学的部位	原発腫瘍
21	組織型	
22	性状	
35	情報源	例：病院カルテ番号、医師名など

したものに、基本項目が追加されることになるかも知れない。

地域がん登録は、治療の評価には関与していないので、患者の積極的な予後調査を、屢々、実施しているわけではない。しかし、さまざまな特性を持つがん患者の全体の生存率を算出することはできる。生存率は最も明確な結果を示す方法である。これを求めるためには、登録された患者の死亡日の情報を収集しなければならない（第12章参照）。

一部の患者に限定して収集する項目

がん登録で、すべての患者に関するすべての情報を収集するのは、明らかに不可能である。基本情報に、一部の患者から収集した情報を臨時に補足してもよい。これによって実施できなかった研究を行いうる場合もある。対象は、患者の特性、腫瘍の種類、時期などによって定める。従って、例えば黒色腫でない皮膚がんの収集は、たとえば、10年に一度、3年間づつ行うことにしてもよい。また、あるがんについての臨時の研究では、収集期間を限定して、広範にわたるデータ項目を収集することになる。がん患者群のQOL測定研究は、一部の患者にしか受入れられないであろう。

表2. がん登録で用いられている収集情報項目

項目番号(a)	項目	項目番号(a)	項目
患者 同定		23 (21)	治療前の臨床進展度
1 (2)	索引番号	24 (23)	手術、病理組織検査後に定めた進展度
2 (3)	個人識別番号		
3 (4)	氏名		
統計・社会的背景		25 (59)	TNMシステム
4 (5)	性別	26 (60)	遠隔転移の部位
5 (6)	生年月日	27 (20)	多重がん
6 (8)	住所	28 (64)	対称臓器の側性
7 (7)	出生地	治療	
8 (9)	婚姻状態	29 (22)	初回治療
9 (11)	罹患年齢	(65-70)	
10 (52)	国籍	予後	
11 (54)	民族	30	最終確認日
12 (53)	宗教	31 (24)	最終確認時の状況
13 (55、56)	職業および産業	32 (25)	死亡年月日
14 (77)	移住した年	33 (26)	死因
15 (78)	父および母の出生国	(76)	"
腫瘍関連		(84)	"
16 (13)	罹患年月日	34 (83)	死亡場所
17 (17)	診断の根拠	情報源	
18 (81)	個々の検査の診断の確実度	35.1 -	情報源の種類： 死亡票、医師、 検査施設、病院、他
19 (57)	発見方法	35.2 -	実際の情報源： 検査施設名、 病院名、医師名、等
20 (18)	解剖学的部位 (ICD-O)		
21 (19)	組織型 (ICD-O)	35.3 -	日付
22 -	性状		

(a)：番号は、「WHO 標準がん登録ハンドブック」(WHO、1976a)の項目番号

項目の国内及び国際比較性

少なくとも同一国内にある登録室は、がん登録に用いる項目の定義やコードを一致させるべきである。したがって、例えば、人口統計に用いるコード（患者の属する人口集団、職業、住所など）は、疫学的解析で分母となる数値を提供する国勢調査局や統計局で使うコードと同一にするべきである。

がん登録により収集された分子と利用できる分母との間に、国内での比較可能性を保つ必要があることから、完全な国際的標準化は難しい。例えば、米国とオーストラリアでは、移民の人種と母国の定義が異なっているかも知れない。こういう情報項目に関しては、個々の項目の細かな用語に国際比較性がなくても、データの収集方法や率の計算方法によって、国際比較性を達成させ得るのである。国際比較性の詳細と程度とは、このように均一ではない。腫瘍の記述とコーディング（第7章参照）は、中でも最大のものであろう。この章での勧告は、分母集団に用いる国勢調査データ以外の変数の比較性のための基本と考えてもらおうとよい。この勧告は、世界中のがん登録室が現に使用している方法に基づいている。

固定項目と更新しうる項目

固定項目とは、後に新しい情報が得られても修正しない項目のことで、例えば治療前の臨床進行度などがこれにあたる（項目23）。誤情報は訂正しなければならないが、これは誤記載ではない。更新しうる項目とは、後に新しい情報が入ってきた時、変更する項目のことで、がんの診断根拠となった検査方法などはその例である（項目17）。ただし、がん患者のフォローアップ調査情報とは区別する必要がある。この時には、古い情報を新しい情報に書き換えるのではなく、古い情報も新しい情報も日付とともにその患者のデータとして保存していく。例えば、多重がんの場合（項目27）。

個人の同定

明確な個人同定（項目2、3）は、すべてのがん登録で必要不可欠である。同じ患者や腫瘍の重複登録を防ぐことが必要であり、また、予後情報を得たり、レコード・リンケージを行うというような、がん登録のさまざまな機能を活発にすることが必要である。ある地域の場合には、国際的に標準化された項目の情報を集めることよりも、個人同定のための十分な情報を得ることの方が、はるかに重要である。個人同定に必要な項目は、国によってかなり異なるので、患者の同定を特定の項目のみに頼るよりは、十分な情報に基づいた同定を行うことに重点をおくべきである。個人同定番号や社会保障番号は、多くの国に存在しており、これらは患者の同定に非常に役立つであろう。

個人の同定に役立つ、また、がんに関する独立した記述変数としても有用であるような、その他の個人同定指標、たとえば生年月日、性などを以下に記述した。正確な生年月日は、個人同定には最も役立つ項目のひとつで、入手可能であれば、常に記録しておくべきである。おおよその年齢も、がんのパターンを記述するには十分に役立つ。

腫瘍に関する記述

このがん登録の中心ともいえる部分は、解剖学的部位（項目20）、組織型（項目21）、性状（項目22）、多重がん（項目27）、治療前の進展度（項目23-25）、がんの診断根拠（項目17）などの項目を含む。解剖学的部位は、作表に際して最もよく用いられる項目である。国際疾病分類－腫瘍学（ICD-O）に準拠して行ったコーディングは、国際疾病分類の現行版（ICD-9）を用いた解剖学的部位のコーディングとは異なる。腫瘍の分類とコーディングについては、第7章に詳細に記述されている。

治療前のがんの進展度は項目23-25によって記載するが、これは初発診断時と治療時との2つの時期の進展度に関連している。その一つは、一般的には臨床進展度といわれ（項目23）、治療を開始する前に、臨床的に評価した進展度である。これにはTNM分類（項目25）

（Hermanek、Sobinら、1987年）がよく用いられている。もう一つは外科病理学的進行度といわれ（項目24）、初回の手術療法から得られた進展度情報がこれにあたり、リンパ節転移に関する組織学的情報や、あるいは患者が治療前に死亡した場合は、剖検で得られた情報などが含まれる。

患者情報の項目の記載とコーディング

患者情報の各項目について、その定義と関連説明とを加えながら、以下に順を追って記す。各々の項目には、いくつかのカテゴリーや分類がある。コーディングは入力操作であるが

（第8章参照）、項目の記述とともに、ここで示すコードを用いると一層便利がよい。腫瘍のコーディングは複雑であり、第7章に別に取り上げた。コーディングは分類を変更すると、混乱してしまう。Payne（1973）は、次のような実際的な問題点を示した。「国内の、あるいは国際的な疾病分類やコードを立案する委員会が、あまりに頻繁に分類やコードを変更するため、がん登録室やそれに類似した組織は、相当な不都合をこうむっている。このような変更が起こると、登録室は、実際に変更が実施された時からそれに従うか、あるいは、変更に対応させるために、それまでに登録されたすべてのデータのコードを変換するか、いずれかになる。前者の場合には、登録データに厄介な不連続性がいつまでも残ってしまい、長期にわたって分析の複雑さに対処しなければならない。また後者では、変換作業に時間と予算がかかり、その上、系統的な間違いを誘発しやすい。」

勧告するデータ項目には、一連の番号をつけた。表2の()内には、WHOが1976年(a)に提唱した院内がん登録での項目番号および「がん登録とその方法」(MacLennanら、1978)に記載されている項目番号を示した。

個人識別

A. 同定項目

項目1：索引番号

登録室は、患者一人一人に登録番号を付加する。この番号は各患者のすべての記録書類や情報項目に付加される。一人の患者が多重がん(項目27)をもつ場合、すべてのがんに同じ登録(患者)番号をつける。それらの原発腫瘍は、部位(項目20)、組織型(項目21)、罹患日(項目16)などで区別できる。この問題は、項目27で、もう一度とりあげる。腫瘍登録番号より患者登録番号を使用する方が望ましい。なぜならば、患者登録番号を用いた方が、多重がんの分析がしやすいし、また患者の予後調査が簡単になるからである。広く用いられている番号の付け方は、最初に登録された年の下2桁の後に、その年の通し番号をつける方法である。例えば、8700001は、1987年に1番目に登録された患者を意味する。1987年の2番目に登録された患者の番号は8700002となる。

登録年は、患者が最初入院、診断された年とは異なるかも知れない。例えば、ある患者が入院、診断されたのが1986年10月としても、1987年の1月まで登録されなかったかもしれない。この場合、登録番号は87から始まるが、罹患の集計年は、罹患日(項目16)によるので、1986年となる。

このことは、第8章でさらに詳しく述べる。

項目2：個人同定番号

多くの国は、個人特有の同定番号を用いている。性や生年月日が、その番号の一部に使われていることもある。個人識別番号を持たない国もあれば、一人2つ以上の番号を持つ国もある。例えば、北欧諸国、マレーシア、シンガポールなどには国民背番号があり、米国には社会保障番号がある。

病歴では、このような個人識別番号の使われ方は一様ではない。どちらかと言えば、治療や入院あるいは患者の保険給付などに関連した管理目的で使用されることの方が多い。ごく一部の患者にしか、個人識別番号として適当な番号が使えない場合は、その番号に頼るわけにはいかないのであるが、それでもとにかく個人識別番号が存在している場合には、それをすべてのカルテのなかに含めてゆくこと(入院時が望ましい)を登録室が病院に働きかけてゆくべきであろう。個人同定番号が広く行き渡っている国であれば、それを索引番号(項目1)として用いることもできる。

チェックディジット（検査数字）も含めて、完全な番号を得るようにする。手書きの番号は間違っているかもしれないことに注意すること（番号の数字の順序の入れかわりや、他人の番号が記入されていることは数多い）。

項目3：氏名

がん登録では、個人同定のために、フルネーム（省略型でない）が必要不可欠である。この項目の情報をえることは一見簡単に見えるが、特に発展途上国では、氏名に多くの問題点がある。出来れば氏名は、同定カードからコピーすることが、常に望ましい。

氏名のつづり：同じ発音でも、つづりの違う名は、屢々ある（例：ReidとRead、PetersenとPedersen）。発展途上国における文字のない言語や方言に関しては、カルテに使われている音声記号（英語、フランス語、スペイン語など）では、微妙な音の違いは表現できていないかもしれないし、また同じ名が、時によって異なるつづりで書かれていることもあるかもしれない。このようなつづりの不明瞭さは、例えばニューヨーク州同定情報システム（NYSIIS）（付録3C中に引用されている）のような、特殊なコーディング法を用いることにより、軽減することができる。

省略形：常にというわけではないが、時々名前の省略形を使っていることがある（例：JamesをJimやJimmyに、RobertをRob、Bob、Bobbyに）。

称号：いつもあるわけではないが、時に称号は個人同定の助けになる（例：Doctor、一部の宗教で用いられるFather、Mother、Brother、Sister、英語圏で用いられるMrs、Miss、Msなど）。

氏名の変更：氏名変更は、がん登録の処理を複雑にする。最も多いのは、結婚に伴う女性の苗字の変更である。また非工業国では、それ以外にも名前の変わる時期がある。発展途上国では、入籍の場合も含め、名前がどんどん追加されていく場合もある（例えば父の名が加わるとか。南米では、母の苗字が加わっている）。このため、複数の氏名を記録する欄を設けておく必要がある（特に女性の場合は旧姓の欄も必要）。

氏名の順序：名前の書かれる順は、慣習により変ってくる。西欧では、文書では苗字は、書式により、最初か最後に来るが、日常会話では最後に来る。アジアでは多くの場合、苗字が最初に来る。氏名の順序は、その地域の習慣にあわせながら、各々の登録で、どちらかに統一しておくべきである。

B. 統計・社会的背景に関する項目

項目4：性別

性別は次に来る同定項目で、カルテには必ず記入されているが、それ以外の多くの情報源には、必ずしも記入されていない。性が書かれなくても、名前（苗字でない）から、或は診療要約の言い回しから、ある程度推測できる地域もあるが、それ以外の場合、例えば、病理報告のみをもとにがんの届出がされている場合には、性別の区別は難しい。手術や薬物などにより性転換している場合には、性別のコードは別に定めるべきである。次のようにコードを定めることをおすすめする。

1. 男 2. 女 3. その他 9. 不明

項目5：生年月日

生年月日は、個人同定を助ける上で、非常に重要である。特に、氏名の種類が少ない場合や、他の個人同定情報が欠落している場合に重要である。生年月日に関連した項目「罹患時年齢（項目9）」は、生年月日（それが判明している場合のみ）から計算できる。反対に、生年月日が不明の場合、現在年齢から概略年を推定できる。これは出生コホートを作るのに役立つ。身分証明書に記入されている生年月日は、推測であることもあるが、すべての記録で一貫して用いられているのであれば、個人同定に役立つ。

国際比較のためには、その地域でのみ使われている日付表記法は、国際的に使用されている標準表記法に変換するべきである。また、年月日は、それぞれの欄に、はっきり記入すること。

25	May	1933
Day	Month	Year

国際的慣例となっている記入順序は例示したとおりである（順序を逆にした方が、データ処理には便利であるが、この処理は、電算機を使えば簡単にできる）。日と年は数字で、月は文字で記入する。これは米国から送られてくるデータに起りうる混乱などを避けるためである。

年月日の記入方法とコーディング方法とは異なる。上記の年月日をコーディングすると、次のようになる。

2	5	0	5	1	9	3	3
---	---	---	---	---	---	---	---

項目6：住所

住所は患者同定に有用である。また、登録対象地域の罹患率の集計を行ったり、細かい地

区別の罹患の変動について検討するために、登録された患者がその地域の住民であることを確認するために、住所は必要不可欠である。さらに登録患者の予後調査を行うときにも、住所は必要である。

ここで記録する住所は、患者が普段住んでいるところにするべきで、入院時の住所とは区別しなければならない。身分証明書があれば、普通これに記入されている住所が患者の普段の住所であり、病院でもコピーされている。この住所は、患者が他の地域からこの地域に入り、一時的に親戚の家に滞在しているかどうかを区別するのに役立つ。病院に入院したり、病院から無料の治療を受ける資格を得るために、患者が意図的に、その地域の住所を書いている場合もある。地域によっては、同じ理由で、身分証明書が他人に借用されていることもある。

地域がん登録ではふだん住んでいる住所をコードすべきであり、また利用しうる分母集団人口と同じ地域分類コードを使用する。また国内や地域内の行政的、政治的境界の変更による影響を最低限に抑えるため、最も細かい地区コードを用いるべきである。こうしておけば再編成が行え、境界線や分母集団の定義の変更にもすぐ対応できる（例：市部と郡部との別、英国や米国での郵便番号の変更など）。その地域に在住していない患者の同定は重要で、このような人たちは、罹患や生存に関する解析を行う際には除外すべきである。これをしないと、特にその地域に有名な治療センターがある場合には、罹患や生存のデータに相当の歪みが生じてしまう。例えば、Tata記念病院ではインド中の頭頸部がんを扱っており、この患者をすべて含めると、ボンベイにおけるこれらの部位のがんの重要度が誇張されることになる。

項目7：出生地

出生地も、個人同定に役立つが、さらにがんの病因究明の糸口を与えるであろう。ある環境から別の環境に移住した人々についての研究で、両環境でのがんの罹患が異なっていることを示すことがある。このような移住の効果の研究は、各国間（例えば日本と米国間）でも、あるいは国内でも行われている。国境が変更される時のことを考慮して、できればその国の正確な出生地を記録しておく（例えば1910年当時ドイツの Breslauに生まれ育った人は、現在、ポーランドの Wroclawに住んでいることになる）。移民した年（項目14）は移民研究においては重要である（下記参照）。

使用するコードはその国の人口動態統計で使われているものと一致させる。これには、がん登録室がカバーしている地域内のほか、国内の他の地域や他国のコードも含まれている。後者の例としては、米国の SEERプログラムで用いられている出生地コードリストがあげられる。WHOも国際的に用いるための国番号リストを用意している（付録1参照）。同じことが国籍（項目10）、民族（項目11）、宗教（項目12）などにも当てはまる。

項目8：婚姻状態

これも個人同定に用いられるが、婚姻の状態は罹病中にも変わる可能性があることを覚えておかななくてはならない。この項目は広く収集することができ、高い比率で正確な情報を得ることができるだろう。用いるコードも、人口動態統計で用いられているものと同じものにするのが望ましい。

項目9：罹患時の年齢

これは、罹患した日における年齢を示す（項目16）。多くの人の中には、正確な罹患年齢の得られない、あるいは意図的に不正確に申告している場合が多く存在する。もし生年月日が不明の場合には、年齢から生まれた年（項目5）を推定して、項目5に記録してもよい。年齢の記録は必要不可欠でない場合が多いが、生年月日のチェックに役立てることができる。

年齢はがんの罹患を記述する際大いに関係するが、正確な年齢が必須というわけではない。発展途上国では、多くの方法でおおよその年齢を推定することができる(Higginson, Oettleら、1960)。例えば、結婚した日が何かの行事と重なっていて、その行事の日時ならわかっていることがある。アジアの中国その他では、生まれた年を、十二支を使って動物で表現している。したがって中国人の場合、干支を併記することによって年齢を確認することも可能であろう。

項目10：国籍

多くの場合、国籍は、市民権と同義に用いられ、その人の法律上の国籍として決定される。国籍を申告していない人や二重国籍の人、その他あいまいな人々では、正確な国籍情報を得ることは難しい。国籍は、住所（項目6）や出生地（項目7）とは区別すべきものである。

もし国勢調査での国籍に関する情報が利用できれば、率を計算できるので、がん登録室は、国勢調査局が使っている定義を用いるべきである。

項目11：民族

これは多くのがん登録室では、必須のものである。その民族の社会や文化の違いが、医療施設の利用や早期発見プログラムへの参加の状況に影響を与えるからである。民族は発がん因子への暴露を規定する文化や習慣の違いのよい指標となる。なぜならば、民族が異なると職業、食習慣その他の習慣や風習も異なってくるからである。主要な民族内のサブグループに関する情報も、特に病因究明の手がかりを得るために重要である。例えば、シンガポールでは、同じ中国人のなかでも話す方言の違いによって、がんの特徴が違うということが示されている(Leeら、1988)。

民族特性のなかで必要な情報は、その国の状況によって異なる。民族を同定するための基礎事項として、もともとの出身国や地域、人種、肌の色、言語の種類、宗教、服装や食習慣、

所属部族、あるいはこれらの組み合わせなどがある。さらに「人種」や「先祖」という言葉には、多くの異なった意味が存在する。したがって、それぞれの登録室で用いるがん患者の民族特性の定義や分類は、どの民族を同定したいかによって決定しなくてはならない。これは国によって非常に異なっており、国際的に標準化した基準を推薦することはできない。

説明が困難なことが起こり得るため、もしこの項目を登録する場合は、基礎とする定義を明確に定めておくことが重要である。がん登録室で用いる民族の定義は、国勢調査などに用いられる公式の定義に対応したものにする方がよいが、より詳しいものにしておく必要があるだろう。国勢調査の数字が得られない場合であっても、民族に関する情報は、罹患の百分率割合の解析の際に重要である。

政治的、或いはその他の理由から、民族が偽られている場合には問題がある。例えば、東南アジアのある国の中国人の場合である。彼らは日常の診療録の中では区別されていないため、がん罹患の特徴を調べるときは、特別調査を必要とする。

項目12：宗教

宗教に関する情報を、特別の任意項目として収集するかどうかはその地域の状況、例えば宗教の数、収集が可能かどうか、がんとの関連がありうるか、などによる。宗教が、場合によると、現代医療への対応や受け入れを決定させ、その結果、悪性疾患に関する知識にも影響を与えることがある。ある宗教集団では、女性は医療を受けるのをいやがる傾向があり（特に男性の医者が診察する場合）、真のがんの罹患率が大幅に低く見積もられることもある。

宗教的な信念は直接発がん物質への暴露に影響を及ぼし、あるいは暴露に影響を与える文化的な違いの指標となる。米国での、特にモルモン教徒やセプンスデーアドベンティストなどの宗教集団におけるがん罹患の特徴により、食習慣、その他の生活習慣のがんの危険因子に関連する作業仮説が多数生まれている。

宗教に関する情報は、民族（項目11）の定義のなかに組み入れてもよい。

項目13：職業および産業

職業とは、産業や雇用の状況（雇用者が被雇用者か）に関係なく、従事している仕事（失業、退職の場合は以前の仕事）の種類のことである。例えば、運輸業であろうが、鉱業であろうが、政府であろうが、とにかく「トラックの運転手」という場合が、これにあたる。

産業とは、経済活動を営む人が働いている施設の事業のことである（国際連合、1968）。職業の中には各々の産業に特有のものもある。国際労働機関（ILO）は1969年に、職業の国際標準分類— 1968（ISCO）を刊行し、現在は改訂版を準備中である。国際連合（UN）は、1968年に全経済活動についての国際標準産業分類（ISIC）を刊行した。これらの分類は、元

来、経済目的のために作られたもので、したがってがん研究用には不適當であることが多い。職業分類と職業名辞典 (CODOT) が英国労働省 (DOE) で1972年に刊行されたが、これには、暴露の可能性の高い特定の職業が詳細に記されている。これはまもなく新しい標準分類に変わる予定で、ISCOと互換性のあるものとなろう (Elias, 1989)。

がん登録室への職業に関する情報の報告は、不十分な場合が多い。しばしば、がんに罹患した時の職業が報告されるが、これは、20~30年前の状況とは関係のないことがある。後者の方が病因とはより大きく関係しているであろう。この種類の情報は取り扱いに注意を要するが、しかし、不完全ではあっても、なお価値のある情報である。

地域がん登録は、コホート研究に経済的な予後調査手法を提供することにより、職業がんの危険因子の研究に重要な役割を演じている (第3章参照)。

項目14: 移住した年

移民労働者も含め、移民集団を扱っている登録室にとっては、興味のある項目である。罹患日と移住日との関連により、新しい環境のなかでの居住期間ががんのリスクに与える影響、あるいは、移住年齢によるリスクの差についての研究が可能になる。例えば、イスラエルでは、移住が開始されて間もなく、子宮頸がんの罹患率が上昇したが、これは、真の罹患率の上昇ではなく、診断される率が増加したことによるものであった。

項目15: 父および母の出生国

かなり多くの移民集団をかかえる国では、両親の出生地も重要な項目である。米国では、日系移民の一世と二世とで、がんのリスクに差が見られたという研究から、特に環境要因 (特に食事関連因子) の重要性が評価された。この場合は、民族と出生国が一世と二世とを分類することに役立った。その他の国では、両親の出生地のみで二世を区別することができるだろう。分母には、通常、父の出生国の方が多く用いられている。

C. 腫瘍とその性状に関する項目

項目16: 罹患年月日

罹患日は、はっきり規定することが難しいので、必ずしも、医師から最初の診断を受けた日である必要はない。病院患者については、外来受診のみ行った患者も含め、がんのために病院を最初に訪れた日、あるいは入院した日を罹患日に定めていることもある。いずれにせよ罹患日は、カルテの中で確認できるはっきりした年月日であり、矛盾がなく、世界各地で入手可能な最も信頼性のある年月日である。このような理由から、罹患日は毎年の予後調査記念日ときめられたり、生存率計算の時の出発点として用いたり、罹患率計算の時には、罹患日として使用される。

もし上記の情報を得ることができなければ、その他の日時を用いる必要がある。そこで、その場合に用いるべき方法を、優先順に示すと、以下のようになる。

- (a) 問題となっているがんについて、病院、診療所、または医療施設での初診、または入院年月日。
- (b) 医師によって最初のがんと診断された日、または最初の病理診断の日。地域がん登録では、罹患年月日を定めるために必要な場合のみ、この情報を収集する。
- (c) 死亡日（年のみ採用する）。がんであったことが死亡票で初めて確認され、その患者についてのフォロバック調査が成功しなかった場合。
- (d) 剖検が行われた場合の死亡日。この場合、臨床的には全くがんの存在を疑われず（「がんの疑い」、「悪性の疑い」のように少しでも疑われた場合は含まれない）、剖検で初めてがんであることが確認された場合。

もし初診と治療のための入院との間にずれがある時は、病院での初診日を罹患日とする（診断と治療が両方とも外来で行われている可能性があるため。例：鼻咽頭がん）。もしがんが別の疾患の治療の際に偶然発見された時、例えば、ある慢性疾患のため、外来あるいは入院して治療を行っている間に症状があらわれ、がんが見つかった時は、がんの診断年月日を罹患日とするのが適当である。

問題は、死亡票のみで登録室に登録された患者の場合である。もし登録室が、この患者の詳しい情報をえることに成功しなかった場合は（項目19参照）、原則的には、死亡日を罹患日とする。

項目17：診断の根拠

罹患率の信頼性の評価に最重要な情報として、罹病中に行われたがんの診断の根拠がある。診断の根拠としては、原発部位の最初の組織診、剖検（この時点で初めて組織の報告の得られることもある）などがある。この項目は後に得られた情報により更新していかなくてはならない。

診断の根拠を考えると、がん登録室は、最低限、がんが組織学的に確認されているかどうかの区別をしておく必要がある。組織学的に確認されていないものを除外すると（病理学者や臨床家が、ときどきこれをやりたがる）、価値ある情報を失ってしまうことになる。というのは、組織学的情報が得られるかどうかは、年齢、腫瘍に接近できるかどうか、医療が受けられるかどうか、そして最終的には患者の現代医療への信頼と態度に左右されるからである。

がんの原発部位の生検は、転移部位の生検とは区別すべきである（例えば開腹術で、臍頭部がんの生検と、腹膜の転移した部位の生検）。また細胞診と組織診とを区別しなくてはならない。

がんの臨床診断を組織学的に確定するためには、がん組織の一部を確実にとってこることが重要である。特に内視鏡的切除検査（気管支鏡、胃内視鏡、腹腔鏡など）の際に、がんの部分を見ておりながら、生検用鉗子で腫瘍を取りそこになってしまうこともある。これらの症例も、内視鏡的診断に基づく患者として登録するべきで、組織学的診断がないからといって除外するべきではない。

剖検の結果の解釈とそれに引き続くコーディングには、次のような注意を払わなくてはならない。

- (a) 剖検の報告には、剖検の組織学的診断が含まれているか。
- (b) 剖検は肉眼的に行われており、組織学的観察は生前にのみ行われているか。
- (c) 剖検には、組織学的診断の裏付けがあるか。

コーディングに際しては、診断方法は大きく組織学的か否かの2つのカテゴリーに分類され、それぞれがさらに4つのカテゴリーに分類される。そのカテゴリーを信頼性の低いものから高いものの順に、以下に示した。組織学的診断をもつか否かという原則を保ちながら、診断技術の進歩も考慮して、欄を2桁に拡大できるようにした方がよい。

非組織学的診断根拠

- 1. 臨床診断（視・触診）のみ
- 2. 臨床検査（X線、超音波を含む）
- 3. 試験的手術、試験的剖検
- 4. 特異的な生化学的、免疫学的検査

組織学的診断あり

- 5. 細胞診または血液学的所見
- 6. 転移病巣の組織診
- 7. 原発組織の組織診
- 8. 剖検時の組織診断、あるいは生前の組織学的診断をもつ剖検
- 9. 詳細不明例
- 10. 死亡票のみの者（DCO）－例えば基本項目データ（表1参照）のみを利用する登録室において、もし他の適当なコードがないときに用いる死亡票のみのデータで、全くさかのぼることのできなかった症例（項目35.1）を登録するときは、死亡票によって存在が確認されても後に別の情報から診断が確認できた症例とは区別するべきである。

項目18：個々の検査の診断の確実度

診断の根拠の他に、コードされた診断の確実さを示すようなコードを含めることを推奨する。病理学者でさえも、剖検報告を作成する際に腫瘍の原発巣を特定できなくて、原発の可

能性のある部位を2-3カ所あげることがある。この項目は、記載されている組織診断が疑わしいかどうかを表す指標であり、一方で、専門家によって検討された後の確定診断であることを示す指標でもある。診断の不確実さについては、部位コード（項目20）の「199.9（原発部位不明）」、組織コード（項目21）の「8000（新生物、腫瘍、悪性、がん）」・「9990（組織型不明）」、それに性状コード（項目22）の「1（悪性度不明）」を使うことによって、ある程度は表現できる。

その他、ICD-0とは別に用いてもよい診断精度のコード：

1. 悪性度及び部位ともに不確定のもの
2. 悪性度は不確定であるが、部位は確定しているもの
3. 悪性度は確定しているが、部位が不確定のもの
4. 悪性度及び部位ともに確定しているが、組織型が不確定のもの
5. 再検査により組織型が不確実になったもの
6. 再検査により組織型が確定しているもの
7. 悪性度、部位、組織型ともに確定しているもの

項目19：発見方法

もしがんの症例がどの方法で発見されたのかに関する情報があれば、その地域集団における経年推移の評価がしばしば容易になる。この項目は、診断の根拠（項目17）とは異なり、その症例がどのような方法によって医学的関心を引くようになったかを表している。特に検診の導入により前臨床期の患者が診断されるようになって、罹患率にも影響がでている。このような患者の中には、結局、症状の現われない者もあるのである。剖検によって初めてがんを発見された患者も、同定できるようにしておくべきである。これは、例えば前立腺がんのように、剖検によって発見される数の多いがんで、どの程度罹患率に影響を与えているかを評価するのに役立つからである。

次のようなコードがよいであろう：

1. 検診
2. 偶然の発見（検査、手術時など）
3. 症状があって受診したもの
4. 剖検での偶然の発見
8. その他
9. 不明

項目20：原発部位：解剖学的部位（ICD-0）

原発部位に関する詳細は、記録されているデータの中で最も重要な項目である。これは登

録データの作表の際の主要な軸となる。

カルテからの要約に際して、腫瘍の部位を、できるだけ明細な情報を含んだ言葉、つまり、例えば「肺左上葉原発の悪性新生物」とか「結腸肝弯曲部の悪性腫瘍」とか「転移性肺癌、原発部位不明」などのように、完全な臨床診断名で記載しておかなければならない。この情報は、補足情報が得られたときには更新する。例えば、最後の例の場合、後で原発部位が報告されれば、その時に部位コードを変更する（しかしこの時には罹患日（項目16）は変更しない）。

第7章で詳細にふれているように、がん登録では極力、国際疾病分類－腫瘍学（ICD-0）（WHO、1976b）を用いるのが望ましい。このなかでは、部位コードが原発腫瘍の解剖学的部位に基づいて表されている。

ICD-0では、部位は、腫瘍の性状に関係なくコーディングされている。したがって、良性腫瘍や悪性度のわからない腫瘍は、悪性新生物と同じ部位コードとなる。そのため、性状コード（良性、上皮内、悪性）を補う必要がある。このコードは組織コードの5番目のコードであるが（項目21参照）、もし組織型がコードされていない場合には、別に性状コード（項目22）を立てる必要がある。

項目21：組織型：形態（ICD-0）

がん登録データを報告するのに最もよく用いられるのは腫瘍の解剖学的部位であるが、組織型は、ただ診断の信頼性を表す指標としてのみでなく、それ以上の重要性の認識が増している。以前は、リンパ腫、白血病、黒色腫、絨毛癌の4つは、国際疾病分類（ICD）のなかでも組織学的診断名で登場していた。しかし、組織型は病因や予後と関連していることが多いため、多くの疫学や臨床研究で検討されるべきものとなった。肝の血管肉腫がビニール塩化物の暴露によって起こるといのように、珍しい組織型は、新しい環境発がんの最初の徴候ともいえる。治療の選択や予後の評価も、組織型の影響を受ける。

がん登録室では、病理報告に記載されている組織型を、完全に登録しなければならない。将来見直しのために標本のブロック、或いはスライドへ戻ることを容易にするために、検査標本番号（項目35.2）も登録するように、きめておく方がよい。

組織診断の言い回しは、コーディングの際に問題を引き起こす。診断に対して議論が起こらないと思われるような、普遍的な腫瘍でさえも、用語が大学によって異なっていることがある。病理学者が ICD-0の組織型の用語を用いてもらえれば、大変助かる。組織型のコーディングの詳細については第7章を参照されたい。

標準のICDの解剖学的部位のコードを用いている登録室でも、組織型の表記にはICD-0を用いるべきである。実際には、ICD-9には ICD-0の組織コードも含んでいる（WHO、1977）。

項目22：性状

もし組織型が ICD-0を用いてコーディングされているのなら、5桁の数字は腫瘍の性状を表している（第7章参照）。データベースのなかに組織コードが含まれていない場合（できれば、すべての登録室で組織コードを含めてほしいが）、以下の ICD-0コードを使って、別項目として性状を登録してほしい。

- 0. 良性
 - 1. 良性または悪性の別不詳
 - 境界悪性
 - 2. 上皮内癌
 - 上皮内
 - 非浸潤性
 - 非侵襲性
 - 3. 悪性

第7章で述べられている勧告に留意して、がん登録室では ICD-0にある性状コード6と9は用いない。

項目23：治療前の臨床進展度

項目24：手術、病理組織検査後に定めた進展度

がんの進展度分類には長い伝統がある。様々ながんの進展度を定めることは、治療計画、大まかな予後の提示、治療効果の評価、治療施設との情報交換の促進のために重要である。これらの機能は主に日常診療と関連しているため、院内がん登録室では進展度を注意深く登録していくことが大切である。地域がん登録では、一般的には、登録するすべての患者の進展度情報を正確に一貫して登録していくのは難しい。地域がん登録で取り扱う進展度は、早期診断の程度を示す情報を提供するために（がんの徴候の重要性、或いはがんの早期発見プログラムの結果に対する一般人の認識度の指標として）、或いは地域集団での生存率の比較性を確認するために用いられる（Hanai & Fujimoto, 1985）。

進展度の登録には、臨床進展度（診断時の医師の臨床的判断を反映する）と、最初の治療で手術が行われたときに手術で確認された所見や、治療の前に患者が死亡した場合の剖検報告を臨床進展度に加味した術後病理組織学的進展度の両方が用いられている。実際には、もしすべての情報を登録すると決めておけば、進展度は、治療の際のあらゆる情報に基づいて得られるので、それらを総合して「進展度」を定める。地域がん登録においては、追跡の過程で得られた進展度情報は、重要ではない。

固形腫瘍に対しては、様々な進行度分類が提案されている。例えば婦人科癌の FIGO分類（米国産婦人科学会、1973）や結腸直腸癌の Duke分類のように、ある特定部位に限定されたものもあるが、その他の分類は、すべての腫瘍に適用できる。後者の中で最も詳細なのは TNMシステムで、項目25で述べているが、さらに綿密な分類もいくつかある。例えば、SEERプログラムの「進行度決定用ガイド」（Shambaughら、1977）では、以下のように分類し、コードしている：

0. 上皮内がん
1. 限局
2. 領域：直接隣接臓器浸潤のみ
3. 領域：リンパ節
4. 領域：直接浸潤及び所属リンパ節
5. 領域：NOS
6. 遠隔（非隣接臓器、遠隔リンパ節、転移）
7. 非限局、NOS
8. どれにもあてはまらない
9. 不明（または未記載）

この分類を利用する時には、それぞれの部位について、所属リンパ節、或いは遠隔リンパ節の定義が必要である。この定義は上記の刊行物（Shambaughら、1977）や、米国対がん合同委員会の「進行度分類マニュアル」（Beahrsら、1988）、さらに「TNMシステム」（項目25）などに掲載されている。登録結果の報告には、上記の分類のカテゴリーのいくつかをまとめる必要があり、多くのがん登録室では、以下のような簡素化した「要約進展度」を用いている：

- 上皮内
- 限局
- 領域
- 遠隔

米国対がん合同委員会も、類似した簡易分類を示しているが、このなかでは、限局腫瘍はさらに腫瘍の大きさに基づいて2群（ⅠとⅡ）とに分類されているが、地域がん登録室でこの情報を得るのは困難である。

上に示した一般的な分類は、どれもリンパ腫の進展度分類には適さない。リンパ腫の場合、通常4期に分類されている。リンパ節とリンパ組織との腫瘍の進展度は、以下のように定義されている：

1. 限局：横隔膜上下いずれか一侧の1つのリンパ節領域
2. 領域：横隔膜上下いずれか一侧の2つ以上のリンパ節領域

3. 遠隔1：横隔膜上下両側のリンパ節領域
4. 遠隔2：1つ以上のリンパ節外臓器への播種

項目25：TNMシステム

がんのTNMシステムは、現在多くの部位について、国際的に確立されている（Hermanek & Sobin, 1987）。

TNMシステムには3つの大きな構成要素がある。「T」は、末尾に腫瘍の大きさと直接進展による侵襲との区別を付加した、原発腫瘍の拡がりを表す。「N」は、リンパ節への転移の有無と拡がりの程度を付加した、所属リンパ節転移の状態を表す。「M」は、遠隔転移の有無を表す。病理学的所見など、それぞれの分野での追加特性は、下につけた数字で表されている。

TNMには非常に詳しいカテゴリーが示されているので、たいいていの場合、上述した要約進展度分類（上皮内、限局、領域、遠隔）の中に凝縮することができる。さらに詳しい内容については、Davies（1977）の報告を参照されたい。

地域がん登録室は、多くの情報源から情報を収集しているため、TNM病期の病院間の比較は難しいことを、よく覚えておく。

項目26：遠隔転移の部位

地域がん登録では、それほど大事な項目ではないが、臨床医は、転移の有無を探ることが多い。第7章で述べたように、項目20のICD-0部位コードは、がん登録では、原発部位についてのみしか言及していない。もし転移部位についての情報も収集したければ、ICD-0部位コード数個分を用意しなければならない。しかし、これほど詳細は滅多に要求されることもないので、1桁の簡単なものが望ましい。例を以下に示す：

- | | |
|-----------|--------|
| 0. なし | 5. 脳 |
| 1. 遠隔リンパ節 | 6. 卵巣 |
| 2. 骨 | 7. 皮膚 |
| 3. 肝 | 8. その他 |
| 4. 肺及び胸膜 | 9. 不明 |

項目27：多重がん

「多重がん」という用語には、多くの問題がある。組織学的に同一、或いは異なる2種以上の腫瘍が、同時あるいは異時に、同一臓器内の異なった部位に、或いは異なった臓器に発生する。登録室内の医系コーダーは、それが単一の腫瘍、すなわち単一原発部位の腫瘍が転移したものか、原発部位の異なる腫瘍であるかを判断しなければならない。登録室は、多重がんを分類するための、明確な手順をもたねばならない。がん登録に用いている定義と多重が

んの報告については、第7章で詳しく述べている。

多重がんは、第8章で提案しているように、索引番号（項目1）に2、3などの番号を付加することによって同定できる。これにより、第2がんや第3がんのための特別なスペースを確保しなくてすむが、一方で、第1がんの索引番号を相互に参照する必要がある。代わりに、腫瘍番号を個人識別番号（項目2）に追加する方法もある（第8章参照）。

項目28：対称臓器の側性

肺のように、左右一対になっている臓器では、病側を把握することが治療の選択のために重要である。その他の場合、例えば網膜芽細胞腫、神経芽細胞腫などでは、片側性が両側性かで病因の定義が異なってくる。対になっている臓器の左右差に関するコードを、予め定めおくことが必要である。例を以下に示す：

1. 右
2. 左
3. 両側
9. 不明

D. 治療

項目29：初回治療

地域がん登録では、この項目は最初の診断から4ヶ月以内に開始された最初の治療を表す。治療方法は、施設により、或いは同じ施設内でも時期により、異なるので、できる限り多くの範囲からデータを収集した方がよいだろう。

腫瘍に対する初回治療を受けなかった患者は、同定できるようにしておいた方がよい。というのは、このような患者は、生存率研究や、がんの自然歴の研究に重要であるからである。

地域がん登録では、できる限り以下にあげるような情報－治療の目的の要約－に限定して収集する。すなわち：

- | | |
|--------------|---------|
| 1. 対症療法のみ | 5. 不確実 |
| 2. 姑息的治療のみ | 7. その他 |
| 3. 治癒的治療－不完全 | 8. 治療せず |
| 4. 治癒的治療－完全 | 9. 不明 |

しかし、登録にたずさわっている臨床医の中には、治療の種類と治療開始日も特定すべきであると主張する者も多い。この時は、初期治療の治療の種類として、次のコードを定める：

- | | |
|------------------|-----------|
| 0. 治療せず（または対症療法） | 4. 免疫療法 |
| 1. 外科手術 | 5. ホルモン療法 |

2. 放射線療法

8. その他の治療

3. 化学療法

9. 不明

冷凍治療やレーザー治療をどのようにコーディングするかは、結論に達していない。

いくつかの治療方法が用いられているので、ある一定期間（例えば初回治療開始より4週間以内）に行われた治療については、すべて治療開始年月日とともにコーディングしておく。

E. 予後

項目30：最終確認日

患者の生存の最終確認日は、その後の患者の来院や、かかりつけの医師との接触、あるいは患者本人から直接確認することができる。生存率を算出する場合、この最終確認日が必要になる（第12章参照）。

もし退院年月日などの情報がなければ、登録時には、最終確認日は罹患日（項目16）と等しくなる。そして、その後新しく確認できたつど、新しい日に更新していく。もし患者が死亡した場合には、最終確認日は消去するか、死亡日（項目32）に置き換える。

項目31：最終確認時の状況

地域がん登録室は、患者の生死に関する情報しか得ることができないであろう。さらに詳細な情報を得るには、患者への積極的な予後調査が必要となるが、これはむしろ院内がん登録室の担うべき仕事である。この項目は生存率の算出の際必要不可欠な項目である。以下のようなコードが望ましい：

1. 生存
2. 死亡
8. 転出
9. 不明

院内登録室は、1. 生存について、例えば腫瘍の有無が確認されたかどうか、などの情報をもっていることが望ましい。

項目32：死亡年月日

死亡票や他の情報にさかのぼることを容易にするためにも、死亡日は、年月日のすべてを登録しておくべきである。この項目は、生存率算出に必要である。

項目33：死因（ICD）

2つの選択が可能である。登録室は、最低限、次のコードを用いる。

1. そのがんによる死亡
2. 他の原因による死亡
9. 不明

これにより、第12章で述べるように、正しい生存率の算出が可能となる。また、もし死亡票から、実際の死因を決定する専門員によって決められた死因を得ることができるのなら、その ICDコードを記録してもよいだろう。死因は ICD全般に及ぶものであり、また根底にある原因から死因を決めるための特別な規則を、実際に適応していく必要があるため、コーディングは非常に複雑な作業である。したがって、これには特別な訓練が必要となる。もしがん登録室のスタッフが自らコーディングをしなければならない場合は、その国の人口動態統計担当部局で訓練を受け、定期的にコーディングの精度を点検しておくべきである。地域がん登録室は、死亡したかどうかのみが確認され、死因についての情報のないものも存在する。例えば医師以外によって作成された死亡票はこの例である。もし死亡票データを人口動態統計担当部局から受け取ることができた場合、既に ICDによる死因コードが与えられている場合もありうる。

この項目が、どの症例が死亡票によって最初のがん登録に届けられたか（項目19）とか、死亡票のみのデータ（項目17または35.1）として登録されたかを定めるためのものではないということを注意しておきたい。

項目34：死亡場所

この情報は、院内がん登録室においても、地域がん登録室においても、役に立つ。コードはどのようなものでもよいが、自宅で、ホスピスで、病院で、というように、その地域の実情に合わせて、登録室が決めなければならない。

地域がん登録では、医療のある種の側面を表すために、この情報を用いることもある。たとえば、病院統計での病院死亡患者数を減らすために終末患者を退院させる傾向など。

情報源

がん登録室は、登録室が用いるすべての情報を組み入れた包括的なコード体系をもつことが必要である（第5章参照）。例えば、病院からの届出患者の場合には病院コード、入院・退院年月日、カルテ番号をも含めることになり、また、検査施設からの届出患者については、検査施設コード、生検日（受付日）、生検番号または標本番号なども含めたコードシステム

となる。

病院のカルテ番号は、がん登録室にない追加情報を得るために、カルテと照会したりすることを容易にする。病院内でカルテが別々に保管されている場合は、その科を確かめ、科別にコードをする。

同じ患者について、いくつかの病院から報告されたときは、地域がん登録室や、いくつかの病院を取り扱っている院内がん登録室では、カルテ番号の他に、病院コードも付加しておかなくてはならない。

「死亡票のみ」(DCO)の患者とは、死亡票によって最初に届出され、死亡診断書の作成にあたった病院や医師にさかのぼっても、死亡票より詳しい情報が得られなかった症例のことである。これらの症例は、登録室のデータを分析する際には、別にした方がよい。第9章で述べるように、DCO症例の割合はがん登録の精度管理に役立つ指標となる。もしできれば、死亡票によって最初に届出されたがん症例についても登録しておいた方がよい。第9章で述べたように、この症例の割合は、登録室の質的管理の程度を示すよい指標である。もし可能であれば、「はじめにがん死亡票から登録したがん患者」の記録を残しておくことよい。このような患者の割合(%)は、症例発見方法の妥当性の有用な指標となる(第5章)。ただし、これらの症例は、上記のDCO症例(項目17)とは異なるので混乱しないようにしてほしい。

これらの項目は、また、情報源を明記することにより、登録室の管理面で役立てることができる。次にあげた各項目を、情報源として推奨するが、各登録室は、自身の方式を作りあげねばならない。

項目35.1 情報源の種類

病院、検査施設、開業医、DCO、その他。

項目35.2 実際の情報源

検査施設、病院、医師の具体名。検査標本番号。

項目35.3 日付

(翻訳：井上眞奈美、花井 彩)

第7章 新生物の分類とコーディング

C. S. Muir and C. Percy

はじめに

新生物の分類は、方法や規則による種類の整理、又は区分を含む。新生物は、いろいろな分類の仕方が考えられるが、臨床医やがん登録にとっても、情報の中の最も重要なものは次の2つの項目である。腫瘍の身体内の位置(同義語:解剖学的位置 anatomical location、部位site、局在部位 topography)と、形態すなわち顕微鏡で調べたときの腫瘍像(同義語:組織学 histology、細胞学 cytology)である。後者は、腫瘍の性状をも示す(悪性、良性、上皮内癌、詳細不明)。がん登録は、登録患者自身の事項を記録することはもちろん、少なくとも、腫瘍をその局在部位、形態、性状により分類するよう努める。

確かな分類には、一致した分類学的名称—系統だった一連の名称—が必要である。それによって、例えば、すべての病理組織学者が、ある特定の顕微鏡下の組織像に同じ名称を与えることに同意する。

分類の責任者は3重の仕事を持つことになる。第一は、概念の変化や使用者の必要性に応じて、分類を改訂していくことを保証すること。でなければ、その分類は使われなくなる。第二に、一時的な変化を、用語や概念に含めないことを保証すること。第三に、定期的に継続して改訂することを保証することである。

便宜上、たいいていの分類は、その構成要素に対して数字を対応させている。それによって、しばしば複雑な一連の情報を数字の形で蓄え、保存し、引き出すことができる。電子、コンピュータ技術の絶え間ない進歩によって今日では、手作業によるコーディングをやめて、コンピュータにコード番号を割り当てさせることによって、記載を直接入力することができるようになった。しかし、この付加された便利さは、疾病分類の基本概念に影響しない。

一般に使われている分類とコード体系は非論理的で、不必要に複雑であるように見える。これは、一つには、「がん」はたくさんの疾病の中の一つでしかなく、それゆえに、長年整備されてきた、より大きな分類体系の一部に割り当てられていることによる。疾病分類の最も重要なマニュアルは、世界保健機構(WHO)によって出版された国際疾病分類(ICD)で、その第9回改訂版が現在使われている(WHO、1977)。詳細は以下に述べるが、現在の様式や構造を説明するのに役立つので、まず、今まで使用されてきた分類の発展の知識を得ることが有用である。

新生物の部位、組織分類の歴史的沿革 (1948-1985)

1948年以前の疾病分類に関する歴史は、ICD-7 (WHO, 1957) の序文に非常によく述べられている。第二次世界大戦の後、国連が設立されて、健康を扱う国連の専門機関としてWHOが設置された。そして国際死因分類に関する責任を引き継いだ。1948年、WHOは、ICDの第6回改訂版 (WHO, 1948) を出版し、その後、10年毎に分類は改訂されている (図1 参照)。

ICDの第二章は、新生物を扱っており、黒色腫、リンパ腫、白血病のような少数の組織型をのぞいて、腫瘍の解剖学的部位に従って整えられた部位分類である。基本的に、新生物の章の構造は、過去40年間変わっていない。新生物は140から239までの100個の連続した3桁の数字に割り当てられている。これらの数字はまた一般に、範疇または項目と呼ばれている。

ICD-6以降、ほとんどの臓器 (または範疇) は、さらに詳しい解剖学的部位を示す 4桁目の数字を用いて、細分類されてきた。例えば、ICD-7で、141.0は、舌基底の悪性新生物に割り当てられている。臓器は、臓器体系によってまとめられた。例えば、ICD-7で、150-159の項目は、消化器及び腹膜の悪性新生物を表している。

新生物の性状は、悪性、良性、性状不詳に分類された。ICD-9からは、さらに、上皮内新生物、性質不詳の新生物にも分類された。ICD-9の構造を表1の例で示した。

1940年代に、初期のがん登録室は、既に、同一臓器における組織的に異なる腫瘍を区別する必要性を認めていた (Clemmesen, 1965)。しかし、腫瘍の組織学的分類を ICD-6は備えていなかった。それは、例えば、肺の扁平上皮癌と肺の腺癌を区別できないことにみられる。双方とも、肺 (ICD-6 162) (ICD-9でも162) の悪性新生物に分類される。そこで、1951年にアメリカ対がん協会は、最初の腫瘍用語コードマニュアル (MOTNAC) を出版した。これは、

表1. 第9回国際疾病分類(ICD-9)の第2章新生物の構成 (3桁番号 140-239)

腫瘍の性質	臓器系	臓器 (部位)	臓器 (詳細部位)
悪性 (140-208)	口腔、咽頭 (140-149)		
	消化器系 (150-159)	食道 (150) 胃(151) 小腸(152) 結腸(153) など	肝湾曲(153.0) 横行結腸(153.1) 下行結腸(153.2) など

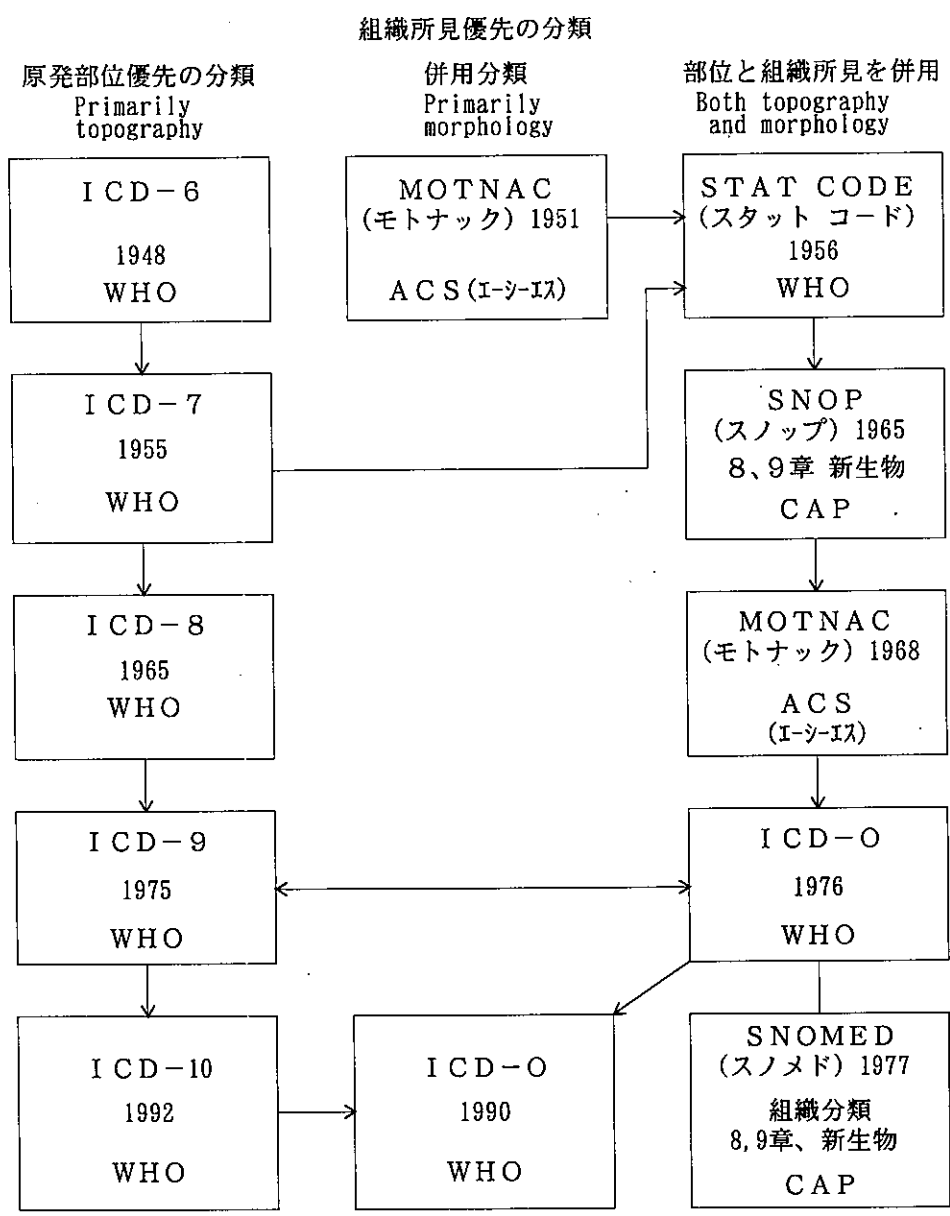


図1. 1948-1985年のコード体系

WHO, World Health Organization; ACS, American Cancer Society; CAP, College of American Pathologists; ICD, International Classification of Diseases; MOTNAC, Manual of Tumor Nomenclature and Coding; STAT, Statistical Code for Human Tumours; SNOP, Systematized Nomenclature of Pathology; SNOMED, Systematized Nomenclature of Medicine.

3桁の形態学コードを持っており、最初の2桁は組織型を、3桁目は腫瘍の性状を表している。当時のがん登録室は、一般に、部位のコーディングにはICD-6の悪性新生物の部分を、形態学コードにはMOTNACを使っていた。この方針は、後に、WHOが1956年に腫瘍統計コードを出版したときに採用された。この部位コードはICD-7 (WHO, 1957) の悪性新生物の章に基づき、形態学コードは性状コードも含めてMOTNACに基づいている (図1参照)。

アメリカ病理学会は1965年、系統的病理学用語集 (SNOP) を出版した。これは、すべての解剖学的部位 (がんの発生部位ではない) をカバーする2桁 (さらに細分類した4桁) の部位コードと、形態学コード (第8、9節が新生物に割り当てられている) からなっている。さらに、病因と機能について4桁のコードがある。アメリカ対がん協会が1968年に改訂版MOTNAC (Percy 他, 1968年) を出版するときに形態学の章にSNOPの第8、9節を用いることが認められた。改訂版MOTNACは、もとの1951年版とはまったく関連がない。しかし、部位についてはICD-8 (WHO, 1967) の悪性新生物の構成に基づき (図1参照)、4桁の形態学コード (性状コードは4桁目) はSNOPから引用されている。

ICDの第9回改訂が進む中、WHOは国際がん研究機関 (IARC) に対して、ジュネーブのWHOのがん部門とICD部門とが協議して、新生物の章 (第2章) の内容と構成について勧告を出すよう要請した。この作業が進む中、新生物に対する論理的で、一貫した、細部にわたる分類の世界的規模での必要性が認められた。このようにして作業班は、部位、形態、性状の3つの軸によって腫瘍を分類しようとする国際疾病分類 (ICD-0) (WHO, 1976) を作成した。部位の章はICD-9の悪性新生物の章に、形態学の分野はMOTNAC (Percy 他, 1968) に基づいている。形態学コードは1桁増え (3桁から4桁)、斜線に続いて性状コードがある。さらに、分化度コード (分化の程度) が形態学コードの6桁目にある。

同じ時期に、アメリカ病理学会は1977年、SNOPを改訂して、系統的医学用語集 (SNOMED) とした。SNOMEDは、ICD-0の形態学の部分を、形態学第8、9節-新生物として取り入れている。一方で、SNOMEDの部位はSNOPと同様、ICD-9 またはICD-0 の部位とは関連性がない。それは、SNOMEDの部位はすべての解剖学的構造をカバーしているが、腫瘍が発生した部位ではないことによる。

分類とコーディング

がん登録は、腫瘍のコーディングにどの分類を使うかを定める時、いくつかの問題に直面する。それは、望ましい詳しさの程度、長期にわたる内部での比較性 (既存の登録室にとっては特に重要な問題である)、登録室間における国際的比較性、などである。

コーディングの根本的原理は、がんを共通の性質によって分類することである。病因、予

後、治療に対する反応性によって分類することは非常に望ましいことであるかもしれないが、そのような情報は、しばしば診断後しばらくしてから得られるものである。現在の知識によれば、腫瘍は、部位、病理組織、性状の3つの軸によって、依然として一番よく表わすことができる。がん登録は、それ故、この3つの軸を用いる国際的に承認されたシステムによって、腫瘍をコードするべきである。このことは、より大まかなカテゴリーで腫瘍分類を行うことを可能にする。

ICD-9は、多くの必要条件をみたしているが、がん登録での使用が勧められる ICD-0の論理性、融通性、組織学的詳細に欠ける。SNOMEDは、ICD-0と多くの利点を共有しているが、ICD分類体系に付された国際的承認に欠ける。SNOMEDの改訂が、その出版元であるアメリカ病理学会によって計画されているが、それ故、以下の頁では、ICDとICD-0だけについて詳しく述べる。

国際疾病分類、1975年改訂版 (ICD-9) (WHO, 1977)

ICD-9は全2巻で出版されている。第1巻は数字のリストで、第2巻はアルファベット索引である。マニュアルは、死亡(死亡診断書)と傷病(病院およびその他の医療機関の診断)の両方のコーディング用にデザインされている。国連条約は、44の国が最新のICDを使って、その国の死亡をコードし、報告するように取り決めている。しかし、条約は、がん登録情報については含めていない。傷病のコーディングに関するいくつかの取り決めが、基本的な死因の選び方に関する取り決めと共に、第1巻の後部に記載されている。

ICD-9では、新生物の章は140から239までの数字で分類している。これらの項目は、次に示すように、新生物の性状によって、さらに6つのグループに分けられている。

分類	グループ
1. 140-199	悪性新生物 (リンパ組織及び造血組織の悪性新生物を除く)
2. 200-208	リンパ組織及び造血組織の悪性新生物
3. 210-229	良性新生物
4. 230-234	上皮内癌
5. 235-238	性質不詳の新生物
6. 239	性質の特定されない新生物

悪性新生物には詳細な解剖学的部位が付されている。多くの3桁分類は4桁目によって、さらに細分類されている。

ICD-9は部位が中心軸であるが、いくつかの形態学的分類を含み、ときには部位と混合している。例えば、皮膚の悪性黒色腫 (ICD-9 172) は、その他の皮膚癌 (ICD-9 173) と区別されている。いくつかの分類項目では、部位によらず、組織が軸になっている。例えば、結合

組織と軟部組織、またはリンパ組織及び造血組織などである。このような「形態学的」区分をもつ悪性新生物の全リストを以下に示す。

結合組織及びその他の軟部組織の悪性新生物：ICD-9 171

皮膚の悪性黒色腫：ICD-9 172

胎盤の悪性新生物（絨毛癌）：ICD-9 181

ホジキン病：ICD-9 201

非ホジキンリンパ腫：ICD-9 200, 202

多発性骨髄腫：ICD-9 203

白血病：ICD-9 204-208

良性新生物（ICD-9 210-229）も、大部分は解剖学的位置によって分類されるが、いくつかの部分は、形態学的であるか、あるいは結合組織またはその他の軟部組織に関連している。すなわち、

脂肪腫：ICD-9 214

結合組織及びその他の軟部組織のその他の良性新生物：ICD-9 215

子宮平滑筋腫：ICD-9 218

血管腫及びリンパ管腫、部位にかかわらず：ICD-9 228

上皮内癌（ICD-9 230-234）の診断は顕微鏡下でのみ可能である。なぜなら、重要な所見は、悪性細胞が上皮組織の基底膜を越えて浸潤していないことであるからである。その様な新生物は部位で分類される。

性質不詳の新生物（ICD-9 235-238）は、明確な組織像を示すが、性質の判定が困難な腫瘍である。例えば、卵巣の顆粒膜細胞腫（ICD-9 236.2）がそれである。

ICD-9の索引は、ICD-0の形態学領域のすべての形態学（組織学）コード（M）をも含んでいる（以下を参照）。

腫瘍学のための国際疾病分類（ICD-0）、初版（WHO, 1976b）

ICD-0は、ICD-9の新生物の章、すなわち第2章の拡張あるいは補足である。

(a) 部位（T）（4桁）

(b) 性状コード（スラッシュに続く1桁）を含む組織学（形態学）（M）（5桁）。前者は、悪性、良性、上皮内、良悪不詳などを示す。

(c) 悪性度（grade I - IV）または分化度（高分化から未分化）の為の1桁。

このように、腫瘍は10桁のコードによって完全に表される。例えば、肺の高分化腺癌のコードは T-162.9 M-8140/31（肺 162.9、腺癌 8140、性状悪性 /3、高分化 1）となる。

部 位

部位分類は、140（口唇）から199（部位不明）までの範囲で、155.2：肝、原発または続発と明示されていないもの、172：皮膚の悪性黒色腫、197：呼吸系及び消化系の続発性悪性新生物、198：その他の特定部位の続発性悪性新生物を除いて、すべてICD-9と同じコード番号を持っている。これらの区分は、性状コード6（転移性）または9（原発、転移の別不詳）、または悪性黒色腫の一つであることを示す形態学コード8720/3-8780/3と結びつけて皮膚の部位173を使うことによって表されるため、ICD-0では使用されない。（ICD-9では、173は「皮膚のその他の悪性新生物」、つまり悪性黒色腫ではないことに留意すること）。

ICD-0は、ICD-9には見られない番号169を含む。これは、ICD-9では200-208に分類される細網内皮系及び造血器系の悪性新生物を、部位としての点から見ている。

- ICD-0 169. - 造血系及び細網内皮系
- 169.0 血液
- .1 骨髄
- .2 脾
- .3 細網内脾系, NOS
- .9 造血系, NOS

組織発生的に、脾は、ここに位置づけられるため、ICD-9での脾159.1はICD-0では落ちてしまう。

ICD-9での項目196（リンパ節の続発性及び詳細不明の悪性新生物）は、ICD-0の部位コードではリンパ節の原発腫瘍のコードに変わった。この番号はICD-0では、ホジキン及び非ホジキンリンパ腫の部位として使われている。臓器に発生したリンパ腫にはその臓器の部位コードが与えられる。胃の悪性リンパ腫はICD-0ではT-151.9 M-9590/39とコードされ、胃腸病学者は胃腫瘍の中に含めている。ICD-9を使うと、この様な腫瘍は202.8とコードされて、リンパ節に発生したリンパ腫と同じになり、原発臓器が示されなくなる。全非ホジキンリンパ腫の20-25%が節外リンパ腫で、リンパ節に発生したものは異なると考えられている。このような新生物を区別して表せることは、ICD-0コードシステムの重要な特徴である。

形態学

病理学者に受け入れられている異なる分類を包含する為に、ICD-0とそれに先立つ MOTONACの著者等は、腫瘍の主要な分類体系に現れるすべての用語に数字コードを割り当てた。例えば、ホジキン病は、一般にはもはや使われていない Jackson-Parker分類(Jackson & Parker, 1944) (M-9660/3からM-9662/3) と Lukes-Collins分類 (Lukes & Collins, 1974) または Rye分類 (Lukes & Butler, 1966) (M-9650/3から9657/3) のいずれにも分類できる。ICD-0初版における非ホジキンリンパ腫の6つの国際分類は、これらの腫瘍に対するコードの使用を

表2. 特定のがんのICDコードとICD-Oコードとの比較

病名	ICD-9	ICD-0 (第1版)	ICD-10	ICD-0 (第2版) (b)
皮膚黒色腫	172	T-173. M-8720/3~ 8780/3	C-43. C-44.	M-8720/3~ M-8790/0
ホジキン病	201	T-196. (a) M-9650/3~ 9662/3	C-81. C-77. (a)	M-9650/3~ M-9667/3
非ホジキン リンパ腫	200, 202	T-196. (a) M-9590/3~ 9642/3 M-9690/3~ 9722/3 M-9740/3~ 9750/3	C-82. ~ C-85. C-77. (a)	M-9590/3~ M-9595/3 M-9670/3~ M-9714/3
多発性 骨髄腫	203	T-169. M-9730/3~ 9731/3	C-90. C-42. 1	M-9731/3~ M-9732/3
白血病	204- 208	T-169. M-9800/3~ 9940/3	C-91. ~ C-95. C-42. 1	M-9800/3~ M-9940/3

a リンパ節外でない場合

b ICD-10とICD-O (2版) 参照

複雑にしているが、一方で幅広い柔軟性を与えている。1982年の Working formulation (National Cancer Institute, 1982; Percy他, 1984) の実施と、ICD-0の第2版 (Percy他, 1990) の中のICDのリンパ腫に関する部分の改訂によって、現在のコード分類が明確になった。上述の点を示すいくつかの例を表2にあげる。

WHOの国際腫瘍組織分類シリーズ (WHO, 1967-1978) は、ICD-0の好ましい用語を選ぶ基礎に使われた。このシリーズ—いわゆるBlue Books—は、最初は、1967年から 1978年間の国際委員会によって作られた。これらの単行本は世界の主な専門家達の意見を代表している。現在26巻のシリーズからなり、各巻は、腫瘍の主要な部位別に出版されている。この本は詳細に図解されていて、カラースライドも買うことができる。最初、コード体系はなかったが、ICD-0の出現によって、関連した形態学コード番号が第22巻より追加された。1978年にWHOは、これらの組織学的区分の要約を用意した。すなわち、このシリーズの最初の20冊 (1967-78) の要約である (Sobin、他1978)。この本では、それぞれの部位で使われている組織学用語とICD-0コード番号を対応させている。これらの分類のいくつかは改訂されつつある。

性 状

これは、形態学コードの5桁目であり、新生物の良性と悪性、上皮内と良悪不詳、原発性と転移性について区別するのに使われる。

コードを以下に示す。

- /0 良性
- /1 良性、悪性の別不詳
境界領域
低悪性能
- /2 上皮内癌
上皮内
非浸潤性
非浸襲性
- /3 悪性、原発
- /6 悪性、転移
二次性（続発）
- /9 悪性、原発、転移の別不詳

悪性度及び分化度

この、形態学コードの6桁目は次の5つに分類される。

1. Grade I (高)分化型
2. Grade II 中等度分化型
3. Grade III 低分化型
4. Grade IV 未分化型、退形成性
9. 悪性度または分化度が未決定、未記載、または適用外

例えば、Grade Iは高分化型というように、適切な分化度コードが各々の Gradeと共に各コードに含まれている。患者の管理に関する臨床医の診断は、腫瘍が高分化あるいは未分化であるという情報で決まることがあるから、このコードは有用である。例えば、婦人科医は、高分化子宮内膜癌（術後照射を伴う、もしくは伴わない子宮全切除術）と未分化型子宮内膜癌（術前照射）とに対して、異なった治療法を選ぶことがある。しかしながら、「組織学的悪性度の用い方は、世界中の病理学者の間で著しく異なっており、多くの場合、日常的には悪性度区分はなされない」（WHO, 1976）。

ICD-0の使用

ICD-0の構成と使用法は、ICD-0の序文に注意深く書かれているので、ここでは繰り返さな

い。ICD-0を使うがん登録は、その協定に習熟することが重要である。

ICD-0のがん登録への適用に関して、重要ないくつかの項目について、特別な困難を経験する項目と共に、以下に概説する。

マトリックスシステム

ICD-0のマトリックスシステムは、分類に関する序文(xix頁)で説明されている。それにもかかわらず、電算機を用いた登録で、プログラムを作成するときに、このマトリックスに関して問題が起こりがちである。可能性として、ほとんどすべての上皮性腫瘍は「上皮内」の時期をもち得るが、ICD-0では、上皮内を含む形態学的分類をもつものは、6つほど掲載されているに過ぎない。性状コード2(つまり上皮内)は、固形がんが上皮内に存在し、上皮内と診断されたならば、いかなる4桁の形態学コードにも続けてコードすることができる。例えば、上皮内乳頭状腺癌は8260/2とコードされる。エラーチェックにかからないように、これらの用語に関してコンピュータプログラムが用意されなければならない。この種の問題は、通常良性であるが病理学者によって悪性とされた腫瘍に関しても起こる。このような場合に注意を促すためにフラッグをたてることは有用であるが、一度診断が確定するとその腫瘍は受け入れて含めなければならない。(逆もまた起こり得る。すなわち、通常悪性であるが良性と診断された場合)。

顕微鏡的確認のない場合

顕微鏡的確認のない腫瘍に形態学の属性を与えることは、勧められない。ICD-0の形態学コードM-9990は、腫瘍が顕微鏡的に確認されていないことを示すコードを望むユーザーの為に用意された。ほとんどすべての登録は、診断が顕微鏡検査に基づいているか、臨床診断か、X線検査等に基づいた診断かを、追加してコードしている。このような項目は一般にがんの診断の根拠と呼ばれている(第6章の項目17参照)。

ICD-0の原発部位と性状コード

一定の腫瘍に対して、性状(悪性、上皮内、良悪不詳)と起源(原発性、転移性、原発続発の別不詳)とを併せて1桁のコードで表すことは、がん登録がICD-0を使用する際に問題となり得る。腫瘍登録は主として腫瘍が発生した局在部位、言い換えれば原発部位、で腫瘍を同定する。そして統計表は、原発部位によって作られるべきである。ICD-0で原発部位を同定する一助として、性状コード/3は悪性、原発を示す。もし、何らかの理由によって、原発部位が不明かつ悪性であることは確かな場合、T-199.9 M-/3とコードされるべきである。(T-199.9は原発部位不明のコードである)。例えば、肺または肝への転移が明らかで原発部位を確定し得ない場合がときどきある。このケースも、原発不明のT-199.9 M-/3とコードさ

れるべきである。

腫瘍登録は、部位不明に割り当てられる症例数を増やしたくはないが、特定の区分が「きれいな」ことが分かる方が、より良い。

ICD-0は、部位特異的形態学用語に備えている。新生物の形態学区分のなかには、部位に特異的なものがある。例えば、腎芽細胞腫（8960/3）と腎臓、基底細胞（8090/3）と皮膚。これらの形態学区分には、適切な部位番号がカッコ内に付け加えられている。この様な形態学区分については、もし診断部位が与えられていない場合、指定部位をコードすることができる。しかし、部位が特定されている場合には、たとえ、それが指定部位と異なっても、その特定された部位をコードするべきである。例えば、浸潤性導管癌には174.-（女性乳房）の部位指定がある。なぜなら、この組織は、通常、乳房に発生するある種のがんに対して使われるからである。しかし、浸潤性導管癌が膵臓に発生した原発がんに対して使われるならば、正しいT番号は157.9（膵臓, NOS）である。

転移巣のコーディング

ICD-0は、一定の臓器における転移の存在をコードするのに、性状コード/6を用意している。しかしこのコードは腫瘍登録では使うべきではない（性状コード/9-原発統発の別不詳-も同じく使用するべきでない）。部位コードは、原発部位のみを指す（上記参照）。

性状コード/6は、例えば、肺または肝組織を受け取った病理学者が、顕微鏡で見て転移性と判断したが、原発部位が分からない場合に使用するために設けられた。病理検査室は、この場合、他の臓器からの肺転移として、T-162.9（肺）とM-/6とコードする。腫瘍登録は、同じ協定に従うことができるであろうが、そうしないことで、原発部位が判明しているが、腫瘍が転移巣に基づいて組織学的に診断されている場合に、このコーディング問題を解決した。例えば、外科医は原発性胃癌から生検するよりも胃の近くのリンパ節を取り除くことを選ぶ場合がある。その様な場合、がん登録は原発部位すなわち胃と、性状コード/3を持つ転移巣の形態学コードをコードするべきである。

もし登録室が、原発がんの顕微鏡検査によって確認された腫瘍と転移巣の組織検査によって確認した腫瘍とを区別したいのであれば、診断の根拠を明らかにする追加コードを使うべきである（第6章、項目17参照）。例えば、もし腫瘍が臨床的に肺の原発がんで、その診断が、転移性扁平上皮癌を示す縦隔リンパ節の顕微鏡的検査によって支持されたなら、T-162.9（肺）、M-8070/3（扁平上皮癌）とコードされるべきである。診断根拠のコードは、この場合 6、すなわち転移巣の組織診であろう。

この様な協定を使うことにより、生検をした転移巣の部位に関する情報は失われてしまう。しかし、遠隔転移巣の部位についての情報を集めている登録では、独立した、遠隔転移部位のコード（第6章、項目26参照）を使用することを勧める。

ICD-9とICD-0の長所と短所

ここでは、ICD-9とICD-0の相対的な長所に関して、様々な観点から議論する。そしてその大部分はICD-10とICD-0の第2版についてもあてはまる。

ICD-9

ICDの主な長所は、それが真に国際的で、死亡統計と殆どどの衛生統計に、全WHO加盟国が使用しているということである。これは、すべての欠点に勝る長所である。しかし、がん登録にとっては、1つのコード番号内での分類軸の組み合わせは、問題を提起する。例えば、ICD-9の172皮膚の悪性黒色腫は3つの軸の情報を伝えている。性状悪性、疾患臓器、組織型の3つである。しかし、皮膚のその他の悪性腫瘍は、ICD-9では173に割り当てられており、4桁目は身体表面のあらゆる部分をコードすることができるが、臨床的にもっと大切な、基底細胞癌と扁平上皮癌の区別をつけることができない。確かに、ICD-9では、大多数の部位において、組織型を区別することは出来ない。ICD-9の索引は、ICD-0の全ての形態学用語を含んでいることが思い起こされる。

それ故、がん登録が、通常のICD-9コードをふり、さらに、ICD-0の形態学コードを付け加える、ということは、完全に実行可能であろう。しかし、そうすることは、組織型を付け加えることで引き出されるメリットの多くを失うことになる。例えば、胃のホジキン氏病は、201 (ホジキン氏病) M 9650/3 (ホジキン氏病) とコードされるであろう。ICD-0では、T151.1、M-9650/3 となり、病変の部位に関する情報が保存される。がん登録においては組織型がコードされることが、必要不可欠である。それ故、ICD-0 を使用すべきである。必要により、ICD-0をICD-9に変換することは、比較的単純である。

専門家の中には、ある解剖学的部位については、ICD-9や、ICD-0で、与えられている局所解剖学的亜(詳細)分類が充分でないと、不満を表明する者がいるが、特別の桁を、特別の研究に限って使用することが示唆される。ICD-8 歯科用改編(WHO, 1978)は、うまく作られた局所解剖拡張のよい例であり、元のICDにまとめてしまうことも可能である。

ICD-0

ICD-0の主なメリットは、コード化と報告書作成のための適切な論理と詳細さを有していることである。詳細さのためにかえって、それを使用することが難しいと思われがちである。反対に、実際の経験によれば、詳細で、また、同意語の索引があるので、正しいコードを与えることが容易で、詳細なコード体系がない場合に生じる判断の必要性を、極力少なくしている。各腫瘍の詳細なコード化が、大まかなコードへの変換テーブル作成の優れた基礎となる。また、小児がんは、概して、局所解剖よりは、組織型によって分類すべきであり、小児がんの国際分類大綱は、ICD-0の形態学と局所解剖学コードに基づいている(Birch &

Marsden, 1987)。

ICD-0 によってコードされたデータの検索と作表は、ICD-9 あるいは ICD-10 の場合より複雑である。しかし、コンピュータで読める形でデータを蓄積している登録室にとっては、このことは余り問題ではない。

ICD-0 は、ICD のように、真に国際的であり、英語、フランス語、ドイツ語、イタリア語、日本語、ポルトガル語、ロシア語、さらに、スペイン語の、8か国語で使用されている。また、広い分野で受け入れられており、病院登録と地域登録の双方で使われている。「5大陸のがん罹患、第5巻」に成績を提供しているおよそ76登録室が、ICD-0 を使用している (Muir他, 1987)。

がん登録室による ICD-0 使用

新しい登録室

新しく作業を始める全てのがん登録室は、ICD-0 を使用することができ、ICD-0 の第2版を使って、局所解剖学と形態学（腫瘍の性状と分化度を含む）の双方を記録すべきである。

既存の登録室

組織分類の有無にかかわらず、ICD あるいは他のコード体系（例えば MOTNAC）を使っている登録室が、ICD-0 への変更を考慮するかもしれない。上述の如く、ICD-0 の詳細さ故に、局所解剖学に関して、長期間にわたり、連続性を損なうことなく維持することが可能である。コンピュータ化した登録室では、登録プログラムの中で、ICDの現在の改訂版への自動コード化のための変換テーブルを組み込んで、ICD-0によるコード化を考慮してもよい。変換に関する詳細な情報は、以下の ICD の変換のテーブルに関する節で述べる。

ICD-10 と ICD-0 第2版

先に述べたように、ICDは、ほぼ10年毎に改訂されている。第10版は、1993年 1月 1日に使用開始となる。常に詳細化が求められ、新しい病気、症候群が認識されてくる中で、ICD-9で使用出来る3桁分類の数は、効果的な拡張のためには不十分であると裁定された。ICD-10 版は、従って、アルファベットと数字の組み合わせを用い、数字のみではない。3桁レベルでは約2000カテゴリーを提供することになる。その内、新生物には、150カテゴリーが割り当てられている。悪性新生物には C00 から C97、上皮内がんには D00-D09、良性の新生物には D10-D36、性質不詳の新生物には D37-D48 が割り当てられた。

既存の4桁の順位は、部位によっては変化している。即ち、結腸については ICD-9の4桁分類の中で、ICD-10で3桁の地位が与えられたものがある。例えば、直腸・S状結腸接合部

(C14)、及び、数個の新しい項目が創設された。注目すべきものとしては、中皮腫 (C45)、カポジ肉腫 (C46)、末梢神経及び自律神経系の悪性腫瘍 (C47)、後腹膜及び腹膜の悪性新生物 (C48)がある。非ホジキンリンパ腫の項は完全に改訂され (C82-C85)、悪性の免疫増殖性疾患 (C88)、及び多重がん (C97)という新しい区分が創設された。ICD-10 は、また、ヒト免疫不全ウイルス疾患 (HIV) に対する一連のコード区分をも与えている。下記に示すこれらの内の1つ (B21) は、がん登録室には特に興味深い。

- B21 悪性新生物を引き起こしたH I V疾患
- B21.0 カポジ肉腫を引き起こしたH I V疾患
- B21.1 バーキットリンパ腫を引き起こしたH I V疾患
- B21.2 他の非ホジキンリンパ腫を引き起こしたH I V疾患
- B21.3 リンパ球、造血組織、及び関連組織の悪性腫瘍を引き起こしたH I V疾患
- B21.7 多重がんを引き起こしたH I V疾患
- B21.8 他の悪性新生物を引き起こしたH I V疾患
- B21.9 非特異的悪性新生物を引き起こしたH I V疾患

ICD-10 では規則上、原死因がコードされるので、死亡統計では、いくつかの悪性新生物が、B21、即ち、新生物以外の章にコードされることになる。死亡情報との照合による「Death clearance」を実施している登録室、あるいは病院退院時診断を検索している登録室は、この見出しによる死亡、あるいは入院の記録を検索する必要がある。B21 の内容から、もしも登録室がその死亡診断書、もしくは症例記録にアクセスしなければ、B21にコードされる新生物の中で、解剖学的部位や性状の情報が失われるものがでてくることは明かである。

ICD-10 の新生物の章の発展と平行して、ICD-0が改訂されたが、注目すべきは、リンパ系、造血器系、及びその関連組織の悪性新生物の分野である。少数の時代遅れとなった用語が削除され、新しい用語、同義語が付け加えられた。胞状奇胎NOS は、初版と同様、良性疾患と考えられる。フォン・レックリングハウゼン氏病を含む神経線維腫症は、骨を除き、性質不詳の新生物に分類される。ICD-10で、これらの用語は、順に、O01.9 と O85 とにコードされる。ICD-0 第2版は、1990年に出版された。ICD-10は、1993年1月1日まで使われないが、WHOは ICD-0 第2版が、局所解剖コード用に、C00-C97 を、改訂形態学コードと共に使うことを許可した。従って、がん登録室は、例えば、1991年1月1日付けからその使用を考えてもよい。

多重がん

一生の中で2つ以上のがんに罹患する可能性のあることは、古くから知られている。数種のがんについて、治療後の生存期間が長くなり、悪性疾患の治療に、それら自身が発がん性を有している化学療法剤が使用されるようになるに従って、現在では全がん患者の約5%が

さらに別の原発がんにかかるかと推定されている。

たいていの登録室は、患者ではなく、腫瘍を数えているので、一人の人が2つ以上のがん罹患していると考えられる条件を定義するための、一連のルールを持つことが大変望ましい。全ての登録室は、それ自身のルールを設定する権利をもっているが、時代的な一貫性と共に、他の登録室のデータとの比較可能性にも注意を払うべきである。国際的な比較の目的のために、IARC は比較的簡単な一連のルールを提案してきた。これらのルールを要約して言えば、次のようである。

- (1) 2つ以上の原発がんがあるかどうかの判定には、それぞれの診断日の間隔は関係ない。
- (2) 原発がんとは、1つの原発部位あるいは組織から発生したがんである。従って、進展したがんでもなければ、再発、転移がんでもない。
- (3) 1つの臓器、あるいは、1対の臓器もしくは組織（ICD の3桁コードによって定義されるような）には、ただ1つの腫瘍しか認められない。（このルールは、ICD-10 が使われだすときに、例えば、骨に関しては、2つの3桁コードに区分されるので、レビューされる必要がある）。
- (4) ルール3は、もしも1つの臓器の複数の腫瘍が、異なる組織型を示す場合には適用されない。表3は悪性新生物の主要な8グループを表示している。特異的な組織型（1及び2、3、5、6、7の番号が付けられたグループ）は、多重がんの定義のためには、異なった組織型と考えられる。一方、グループ4及び8は、組織学的には十分に区別できなかった腫瘍であり、それ故、他のグループと識別する事は出来ない。

IARC は、さらにこの分野に関連して、下記の定義を作成した。

Multifocal (多巣性)：不連続、即ち、同一部位あるいは組織から発生した他の原発がんとは、明らかに連続性がない（例、膀胱）。

Multicentric (多中心性)：リンパ組織もしくは造血組織の異なる部位から発生する原発がん。

上述のルールにより、多巣性腫瘍も多中心性腫瘍も、異なる組織型を持つのでなければ、1回しか計上されない。

上記の定義は、「5大陸のがん罹患」の様な国際的な刊行物に罹患数を報告するときに使用されることが強く推奨される。しかしながら、これらの単純なルールは、臨床研究のためには不十分である。

表3. 多重がんの定義の中で、組織学的に異なると判断するための悪性新生物の群別 (Berg 分類の変法)

I. 癌腫 (Carcinoma)		
1.	A. 扁平上皮癌	805-813 (a)
2.	B. 腺癌	814, 816, 818-823, 825-855, 857, 894
3.	C. その他の特定の癌	803-804, 815, 817, 824, 856, 858-867
4.	D. 不特定の癌 (Carcinoma, Nos)	801-802
5.	II. リンパ腫	959-974
6.	III. 肉腫及びその他の軟部組織	868-871, 880-892, 899, 904-905, 912-934, 937, 949-950, 954-958
7.	IV. その他の特定の(及び特定部位の)型の癌	872-879, 893, 895-898, 900-903, 906-911, 935-936, 938-948, 952-953
8.	V. 特定されていない型の癌	800, 999

注：a, 数字は、ICD-O組織分類の最初の3桁を使用している。

死亡診断書の新生物のコーディング：がん登録との関係

多くのがん登録は、死亡診断書を利用する許可をえている。理想的には、登録は、記載死因に関係なく、すべての死亡票と登録記録とを照合できるべきである。この、いわゆる "death clearance" によって、登録は、生存率を計算し、届出のない癌死亡者を発見することができる。多くの登録室が全死因の死亡診断書利用の許可を持っているが、一方では、がんとコードされたものだけについての情報しか得ていない登録室もある。その場合は、複合死因のコーディングが実施されているのでなければ、原死因と考えられた新生物についてのみ知ることになる。

原死因決定の選択とコーディングの規則は、複雑であり、その解釈が腫瘍のコーディングに影響を与えるので、研究の価値がある。ICD 第10版は、独立した多部位に原発する悪性新生物に対して、新しい区分を作った (C97)。これは、通常なら死亡診断書のコーディングに使用されることになるであろう。本質的にこの区分は、1つ以上の独立した原発性腫瘍の存在に注意を引きつけるが、それらの部位を同定するものではない。一方、ICD-9のコーディング規則では1つの部位の選択を強いており、他の新生物の存在に関する情報は、失われてしまう。がん登録は通常は多重癌の存在を同定できるが、一方、死亡診断書の中の多重癌の扱いには問題が残る。

顧問のアドバイス

登録室に届けられる腫瘍に関する情報は、不完全であるかもしれない。これは、情報の欠落または関連様式への不注意な記載によることがある。推測するよりもむしろ、詳しい情報を提供できる届出人に連絡するよう、あらゆる努力をするべきである。それにもかかわらず、すべての登録室は、むずかしい問題を解決するために、登録で使うコードに慣れた医学上の顧問を持つべきである。例えば、腫瘍が直腸か結腸のどちらに原発したのかを決めるのが、しばしば、困難な場合がある。もし、可能ならば、この顧問がその様な症例をレビューして、判断を下すべきである。もう一つの困難な部位は肝である。登録が、ICDまたはICD-0のどちらを使っても、肝にある癌が原発か続発かの決定を下さなければいけない。もし、続発性であれば、または原発、続発の区別が不明であれば、原発部位には不詳がコードされねばならない。腕、足、または身体の他の部分という様に、不明確な部位の時は、ICD-0の索引部分が助けとなる。組織型は、どの様な組織から腫瘍が発生したかを示す。癌は皮膚に、肉腫は結合組織に、骨肉腫または軟骨肉腫は骨に発生しやすい。もし、これらの用語が見つからない時は、適切な、不明確な部位 195.-を使用する。

検索と報告

もし、データが検索できないならば、コーディングは役に立たない。ICDとICD-0は、どちらも検索に適している。すべての登録室は、少なくとも毎年、データを検索、作表すべきである（詳細は第10章参照）。最小限、部位別、性別の表を、ICDまたはICD-0で使用されているコードに従って作成する。もし、ICD-0がコーディングに使用されているのならば、作表の目的のために、ICDに変換されるべきである。もし、この変換が不可能な場合のみICD-0の部位コードで作表するべきである。そして、これは、種々の組織学的区分を分割する表によって補足されるべきである。ほとんど1,000近い組織型があるので、グループ分けが必要である。これは、共通の区分をリストすることにより、部位と部位との対応で、実行できる。もっともらしい頻度の推計には、Cutler & Young(1975)と Young、他(1981)を参照する。

数年にわたってデータを検索するとき（年次推移）、変換あるいは再度のグループ分けをする必要のある部位が存在することがある。各ICD改訂版-7→8→9→10-は、ある変更を経してきた。使用者は、研究する部位変更を注意深く調べなければならない。コード番号が変わっただけでなく、例えば、乳房はICD-7では170で、ICD-8とICD-9では174（女性）、ICD-10ではC50に変わったが、分類の内容も同様に変わった。例えば、ICD-8では、上皮内の区分は、子宮頸部の上皮内(ICD-8 234.0)ただ1つであった。他の上皮内新生物は悪性腫瘍として数えられた。コードの変更については注意を払うことができる（次節参照）が、内容の変更が与える影響については、評価することが困難である。

ICD 変換テーブル

ICD の新しい分類、新しい改訂が使われ始めると、長期間にわたる成績を報告するために、がん登録室は、これまでの分類でコードされたデータを、新しいコードに変換する必要がある。がん登録室は、ICD-0 に従ってファイルを維持し、結果を、例えば、年報として、さらには「5大陸のがん罹患」シリーズへの掲載のために、ICD-9 を用いて、報告しているかも知れない。米国の国立がん研究所は、Percy によって編集された、新生物のための一連の変換テーブルを作成してきた。最近の変換テーブルは、磁気テープでも、マニュアルに記載されたものとしても、入手できるようになっている。現在入手できるものは、ICD-8からICD-9 (Percy, 1983a)、ICD-9 からICD-8 (Percy, 1983b)、新生物用の ICD-0 からICD-8 (Percy, 1980)、そしてICD-0 から ICD-9 (Percy & van Holten, 1979) への変換テーブルである。

多くの人が ICD-9 からICD-0 への変換を希望する。データは詳細な版から、より簡略な版へと、容易に変換され得る。しかし、逆方向は容易ではない。上述の如く、ICD の用語のほとんどは、局所解剖学的で、悪性腫瘍の形態学については、悪性黒色腫、絨毛上皮がん、軟部組織の新生物、リンパ腫及び白血病を除いて、考慮されていない。局所解剖学コードを変換することは可能であるが、形態学コードを変換することは不可能である。例えば、ICD-9 の 162.9 肺の悪性腫瘍は、ICD-0 の T-162.9 に変換され得るが、形態学のフィールドは、ICD-0 では、必然的にブランクとなり(-/3)、ICD-9 から ICD-0 への変換は、従って殆んど価値がない。ある改訂から別の改訂へ変換する時、ユーザーは、アルファベット索引に掲載されている用語の多くが、改訂毎に違ってコードが指示される場合があるということを承知しておく必要がある。もしもこの用語が相当な頻度で使われるのであれば、統計値に影響を与えうる。この具体例は、神経芽細胞腫である。即ち、この用語は、もしも部位についての記載がなかったなら、ICD-8 においては 192.5 -交感神経系 にコードするように指示された。一方、ICD-9 では 194.0 -副腎にコードするように指示されている。この結果、ICD-9 が使用され始めたとき、副腎がんによる死亡の見かけの大きな増加となって現れた。

時を超えての罹患データの比較性は、がん登録の大切な機能であるから、登録室の中には2つの異なる分類体系でがんをコード化してきた所がある(例えばアイスランドとデンマーク)。これには、登録過程の中で、電算機が広範囲に使用されるようになったことが大きく影響している。デンマークがん登録室の1943-1977年の期間のデータは、ICD-7 の拡張版に従ってコードされている。1978年以降の罹患例は ICD-0 でコードされ、さらにコンピュータによる変換テーブルが、自動的に相当する ICD-7 のコードをふっている。このことにより、40年を越える期間の罹患率を、比較性を保った上で直接作表することが可能となっている。

ICD の改訂

通常 ICDの改訂は10年毎に実施されてきたが、ICD-10版については、改訂が広範囲なので、ICD-9の期間を15年に延長することが、WHO 加盟国によって決定された。

ICDを周期的に改訂していくことは、がん登録室にとって問題を生じる（そして、保健統計の他のユーザーや提供者にとっても）。なぜなら、注意深く改訂されなければ、長期間にわたるデータを比較することが困難になるからである。時間的連続性の諸問題に配慮がなされておれば、通常はより詳しくなっている最新版から、旧版へ変換することは、情報を縮小することにより（下記も参照）、常に可能なはずである。改訂は、全ての統計システムの仕事量を増加する。というのは、新しいコンピュータプログラム、照合のシステムを書き換える必要があるし、出力表を工夫しなくてはいけないし、さらには、一連のコード番号を既に学んでいる登録室のスタッフが、新しいコードを学ばないといけないからであり、これには作業の遅れとミスを伴う。

提案されている変更を、実際に取り入れる前に、フィールド調査により評価しておくことが、極めて大切である。ある版の使用期間が延長されるに従い、誤りの修正には長期間かかるようになるからである。この意味で、ICD-0 第2版ではフィールド調査が広範囲に実施された。

（翻訳：北川貴子、津熊秀明）

第8章 手作業及び電算機によるがん登録

R. G. Skeet

序論

表面的にみると、1年間に数百の新発生例を扱い、書類整理箱に詳細情報を格納している小規模の手作業によるがん登録と、大規模で、ほとんど電算機の端末 (VDU) だけで構成されているような、高度に電算化されたがん登録との間には、恐らく共通点がほとんど無いように見えるかも知れない。これら2つの両極端にある登録の操作手順を1つの章で記述することは、不適當に見えるかも知れない。しかしながら、両者を詳細に検討してみると、機能的構成要素は同一であることがわかる。同じ基礎的作業が各々において実施されなければならない。つまり、異なるのは方法だけである。現代の電算システムの特質は、新来者にとっては現実は何が遂行されたかを理解することが必ずしも容易ではない。症例が登録されたとき (第5章参照)、がん登録室での活動は普遍的である。つまり、活動は、主としてデータを表にし、解析するために準備をすることと関連している。この章では、手作業による仕事と、種々の利用可能な電算機による解決法との、両者を記載するために、手作業及び電算化された登録の両方の操作手順を機能毎に概説する。1つもしくは複数の、ひな型となる登録の手法の記述というよりも、がん登録に適用される概念と原理とに、より大きな関心がある。4つの異なった地域がん登録室の操作手順を、付録3に記述した。

登録室の行うべき作業

全てのがん登録室は、方法の如何を問わず、下記の作業を実行しなければならない。各々の登録室に入る情報源の量は、しばしば登録室以外の多くの要因に依存している。

データの収集

データ収集のための何らかの仕組みがなければ、如何なる登録室も成り立たない。データ収集については、第5章で詳しく述べた。

レコードリンケイジ

登録室は、例えば、病院、病理検査所や人口動態局など、複数の情報源から、しばしば同一患者に関連した記録を受け取る。これらの記録は全て、各患者の詳細が揃うように、また、同一腫瘍について重複登録がないように、同一患者に連結されなければならない。レコードリンケイジは極めて重要な作業であり、その大切さを強調しておく。

データの組織化

科学的研究のためのデータは、きちっと整頓され、維持されなければならない。情報は、登録室には多少とも組織立てられた様式で、一部にはがん登録のために特別に創られ、うまくデザインされた様式で、一部にはより記述的で、主には他の目的のためにデザインされた報告書の形で、入ってくる。電算化したデータが、既に処理された形で、あるいは部分的に処理された形で、到着することもあるが、データのさらなる組織化が、登録室ではなお必要とされる。

媒体変換

手作業による登録室でも、情報がもとの記録のまま完全に維持されるということは、殆どありえない。電算化された登録室では、紙面上の情報が、機械で読める媒体、パンチカード、磁気テープあるいは磁気ディスクに変換されなければならない。電算化された登録室では、データを複数の媒体で保持していることもある。

問い合わせの発生とフォローアップ

ある項目の情報の収集により、登録室が既に得ている情報が不完全であったり、不正確であると、気がつく場合がしばしばある。例えば、最近死亡した悪性疾患の記載のある死亡票の患者が未登録の場合は、がん登録室がより早期の段階で情報を入手し損ねた可能性を示している。その場合、登録室は詳細情報を得ようとして、あるいは何らかの不一致を解決しようとして、さらに問い合わせをしなければならない。多くの登録室は、能動的であれ、受動的であれ、登録患者のフォローアップを最も重要な機能の1つであると考えている。能動的フォローアップは、登録患者に関する追跡データの定期的請求を包含している。

データ解析

がん登録データの解析は、第10-12章で扱われるが、完結させるため、そして、もしもこの最終段階の作業を行わないのなら、それまでの仕事が無益になるということを強調するために、ここでも触れておく。

上記で記述した各プロセスについて、以下に順を追って論じる。適切などころでは詳細に述べるが、がん登録室によって処理方法が相当異なるので、関連する主要原理に重点をおいて記述する。各登録室で遭遇する同様の諸問題に対して、解決方法は唯一ではないので、作業が絶対的な条件で、どの様になされるべきかを記述するつもりはない。

レコードリンケイジ

複数の報告

レコードリンケイジの問題を考える前に、重複届出、重複腫瘍、及び重複登録の、基礎概念を理解しておくことが重要である。

重複届出

1人のがん患者における1つの腫瘍に関して受け取った、複数の報告を指す。もしもある患者がある病院でがんと診断され、治療のために別の病院に紹介される場合、両病院とも、その患者をがん登録に通報することがよくある。登録室は、これらの報告を重複届出と認識できなければならない。

多重がん

時として、一人のがん患者が複数の原発腫瘍に罹患するが、その各々について、独立して登録するのが慣例である。なぜなら、がん登録は、現実にはがん患者数というより、むしろ原発腫瘍の数を数え上げているからである。登録室は、原発がんの過剰及び過小登録を避けるために、重複がんを構成するものが何かについて、明確な定義をもつことが重要である。重複がんの研究は、それ自体重要である（第3章参照）。重複腫瘍に関して、国際的使用に適した一つの定義が第7章に提示されている。

重複登録

重複登録は、がん登録が1つの腫瘍を複数回計上することで、レコードリンケイジのプロセスの失敗の結果として発生する。

リンケイジのプロセス

がん登録におけるレコードリンケイジの目的は、同一人に属する各記録を集めて、その記録が、既に登録されている腫瘍（症例）に関するものか、あるいは新発生の原発腫瘍に関するものかを決定することである。

最も単純な形で、このことが図1で示されている。はじめに、患者が未登録と判明すれば、その場合は新規登録が行われる。既登録であれば、もしも届出が、既登録の腫瘍と異なった原発腫瘍に関するものであれば、その時のみ、その腫瘍について新規登録が行われる。手作業によるリンケイジでも、電算化されたリンケイジシステムでも、同一の基本的プロセスを踏むことになる。新規患者にせよ、新規腫瘍にせよ、新規登録が行われるとき、登録室は1つの番号を発行する。この番号は通常承認番号、もしくは登録番号と呼ばれる。

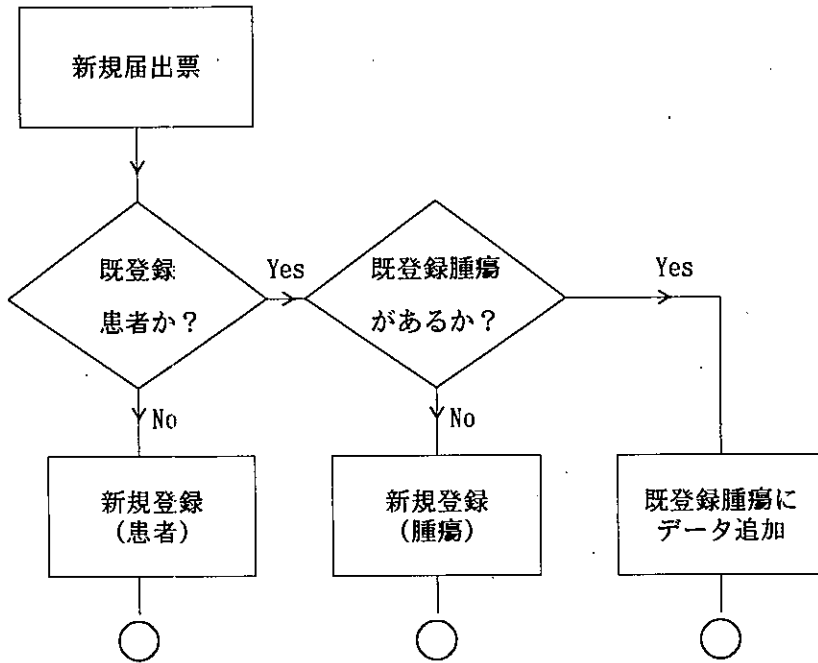


図1. がん登録室における情報のリンケイジの基本的な手順

名前	性	生年月日
住所	病院番号 1 2 3 4	
診断 1 2 3 4	登録年月日	
	索引(登録)番号	

図2. 患者索引カード(見本)

手作業によるリンケイジ

登録簿を氏名リストのような形に編纂する目的は、新規患者が既登録かどうかを確認するべく、氏名リストと照らし合わせが出来るようにする為である。個人同定番号が使用されている国を除けば、新規届出患者の名前を既登録の名簿と照らし合わせることが、この作業を実行する上で、用いる唯一の方法である。

名前の識別力は限られているので、名前のみでは通常不十分である。非常にありふれた名前の場合には、識別力はとりわけ低い。多くの登録室では、誕生日もリンケイジに用いている。というのは、誕生日は識別力を随分増すからである。もしも、名前と誕生日が間違いなく正確に記録されるとしたら、正確なレコードリンケイジのために、他の指標が必要になるとは考えられない。しかし、そのようにはなっておらず、たいていの登録室では、患者住所と、恐らくは結婚前の姓をも、リンケイジの精度をあげるために、用いている。

レコードリンケイジの伝統的手法は、図2で示す例と似た患者索引カードのファイルを維持することである。がん登録室に来る全部の新規記録が、この索引カードと照合され、その結果として、マッチするものが見つかるかどうかで、2群に分けられる。マッチするものがない場合には、患者に新規に登録番号（以下を参照）を与えて、新規患者索引カードが作成され、索引カードファイルに加えられる。

一般的に、このプロセスは、通常毎日、あるいは毎週、毎月のバッチ (batch) 処理で実行される。入ってくる様式は、氏名のアルファベット順にソート (分類) され、それから索引ファイルと照らし合わせられる。その場合、恐らくは数人の事務員で仕事を分割し、各々は索引カードファイルの異なる部分を使う。様式が、あるいは、誕生日でソートされることもある。

新規に登録すべき症例が同定され、番号が付けられた後、患者索引カードがタイプされ、ファイルされる。このファイル作成プロセスは、実際には索引カードの2回目の検索に当たる。それ故、この作業を別の事務員が実施するようにするとよい。例えば、一人の事務員がアルファベットの前半を検索し、もう一人が後半を検索する場合、患者索引ファイル作成プロセスには、役割を反転させるべきである。不注意に見逃したペアがこの方法で検出され、誤りが修正される。

次のバッチ (1束) のチェックが始まるまでに、患者索引カードのファイル化が終了していることが絶対に必要である。同一患者に関する重複届出が、しばしばほんの僅かの期間内に登録室に届く。そのため、届出票が既に登録室に届いており、索引カードのファイル化を待っている患者を、索引ファイルから検索することに大変な無駄が費やされる。

名前が書類によってまちまちに綴られる為に、事務員が大きな苦勞をする。著者自身の名前 Skeetは、この点である種の記録の適例であるかも知れない。すなわち、Skeat、Skete、Skate、Sheet、Streetが、それらの複数形とともに、聞き間違い、あるいは転記間違いの結

果として、しばしば使われる。

誤りが名前のはじめに近ければ近いほど、重複登録が発生するチャンスが大きくなる。このことはローマ字のアルファベットの使用者に限った問題ではない。唯一満足すべき解決法は、純粋なアルファベットシステムと表音式綴との、ある種の折衷様式で、索引カードをファイルすることである。この場合、似た発音の名前と一緒にファイル化される。かくして、Symonds、Simmons、Simons、Symon、及び Simon は、索引ファイルの中では、全て一緒に、恐らくは 'Simmons' の下に、ファイル化される。事務員が正しく検索できるように、他の綴の適当な箇所に、ガイドカードが挿入される。手作業による索引の検索は、経験が本質的役割を演じる相当な技術に発展し得る。経験のある事務員が、登録室の責任者が見い出せない名前を、索引から見い出すことは、よく知られている。

名前の使用は、地方文化の産物である。結婚に伴って妻が夫の姓を名乗ることは、多くの文化で共通しているが、そうした文化でも例外が起こり得、また増加しているように見える。名前は他の理由で変わるかも知れないし、短縮やニックネームも使われることがある。これが重大な問題となる所では、組み合わせ索引の作成が役立つかも知れない。しかし、いずれにせよ、登録室は患者索引ファイルを地方の習慣に合うように構成すべきである。名前は決してアルファベット順に並べた索引から除いてはならない。もしも患者が名前を変えた場合には、新旧双方の名前を索引ファイルに維持すべきである。これは、同一人物について、もう1枚カードを作成し、検索者に元のカードを参照させることを意味する（ときどき also-known-as、あるいは 'aka' card と呼ばれる）。重複がんをもつ患者については、全診断を記載した1つの索引カードを持つか、それとも、それぞれのがんについてカードを1枚ずつ作成し、便利のために、ホチキスで留めておく。

患者索引が大きくなるにつれて、その中で死亡例の占める割合が大きくなる。確かに、死亡日から3年も経って、さらに新たな報告が届くことは殆んど有り得ないのであるから、これらの患者の索引カードを主要患者索引から、使用頻度の少ない補助死亡ファイルへ移すことを考慮すべきである。これには多くの作業を伴うが、大規模の登録では、主要オフィススペースが特別に貴重であり、保管スペースのみからいっても、この分離は非常に重要になるかも知れない。基本的に現役でなくなったものを主要索引から取り除くことにより、ずっと小さなファイルにすることができる。こうしたファイルは使いやすしいし、誤りの頻度も少なくなる。

電算機によるリンケイジ

がん登録では、電算機によるレコードリンケイジに、2通りの方式があり、大きくわけて、オフラインとオンラインの範疇に分かれる。

オフライン・レコードリンケイジは、準備された記録のバッチ (batch) を、ディスクもしくは磁気テープで電算機に送ることからはじまる。そして電算機が、新規記録の個人同定情報を、システムに既に存在している記録と突き合わせをする。各新規記録のマッチングの程度を示すために色々な方法が用いられる。これには、氏名の完全一致が、近似マッチより、高スコアを獲得し、一方、存在する名前と何のマッチもない場合に、ゼロスコアを与えるような、ある種のスコア方式を含む。同様のスコアが誕生日の一致度についても計算される、その他のデータ項目が比較の目的で使われ、それぞれについてスコアが計算され、そして最終的な重み付きスコアが計算される。スコアが、それから評価される。すなわち、ある臨界レベル以上はマッチが正しい (同一人*) と判定され、別のあるレベル以下のものは、マッチしない (異なる人*) と判定される。これら両スコアの間にくるペアは、マッチが有り得る (同一人の場合もありうる*) と判定される。これらについては、通常手作業による検査のため、全データが出力され、その結果、事務員が個々に決定できるようになる。これらの方式は、データ処理に際して、これ以上の手作業が不必要な、例えば、中央登録室にくる病院からの電算ファイルを扱う場合に、特に有用である。マッチした記録は、自動的に既存の記録を更新し、不一致の記録は、新規に登録される。このタイプの登録体系は、オンタリオ州腫瘍登録室で使われており、付録3(C)で詳しく紹介されている。

オンライン・レコードリンケイジは、紙面上の記録が VDU を使って電算システムに入力される場合に、最も有益である。データそのものの入力に先立って、オペレータは、リンケイジに必要な氏名、誕生日、その他の情報を打ち込む。そうすると、電算機は、同一もしくは類似の情報を持つ症例を、ファイルの中から検索し、スクリーン上に、同一の可能性のあるものを出力する。これらを基にして、オペレータは、症例が既登録か、新規登録になるものかを判断する。これらのシステムは、色々な表音方式を使って、可能性のある同一人物を同定するための、精巧な方法を持ち得る。また、検索が異なった様式をとり得るので、非常に柔軟性に富んでいる。例えば、誕生日を固定させて、一定の文字列で始まる全登録例の名前を表示させる事が出来る。あるいは、名前を一定にして、誕生日にかかわらず、全症例を表示する事も可能である。

このようなレコードリンケイジは、注意深く設計された索引を使うことにより、1レコードにアクセスするスピードは、システムに登録されている症例数に直接には関係せず、極めてすばやくなされ得る。非常に複雑かも知れないが、うまく組み立てられた索引を用いることにより、1秒以下で100万件を越えるレコードのファイルから正確なマッチを見つけ出すことが可能となる。このことは、データベース管理技術を使う高級な電算プログラムと、大規模な登録の場合には、永久的にオンラインである大容量ディスク記憶装置を持つことと、依存している。

* 訳者が追加した

手作業によるリンケイジとは大変違って見えるが、原理は全く同じである。索引は、カードで保管される代わりに、ディスクで保管され、一方、電算ソフトウェアが、もし存在するとすれば、どの引き出しにカードがあるのかを知っている検索者（人）に、取ってかわっている。索引ファイルは、もちろん電算機自身により維持される。新規症例であると認識されると直ちに、電算機は自動的に索引記録を作成する。こうして、バッチで処理する必要性がなくなる。名前あるいは誕生日の修正は、元の情報を削除する事なく、自動的に索引にフィードバックされる。また、入ってくる情報をアルファベット順にソートする必要もなく、死亡例を別のファイルに移す必要もない。

登録番号の付け方

殆どどのシステム、とりわけ電算化されたシステムにおいては、データを、アルファベットよりも、番号で蓄積する方が便利である。リンケイジの結果、新規患者であると認識されれば、患者に患者登録番号が与えられる（第6章参照）。すなわち、最も広く使われているナンバリングシステムでは、始めの2桁は登録年（これは定義されるが；以下を参照）を表わし、次に、その登録年に登録される症例の順序に従ってつけた番号が続く。それ故、1987年の登録日で登録される第1番目の症例は、8700001、2番目は8700002、等の番号が与えられる。登録年は罹患年とは異なることがある。1985年と1986年とに診断された例が、1987年に登録されることがある。罹患を計算する年は、順に1985年及び1986年であるけれども、これらの例では、接頭辞87を割り当てられる。

同一患者における多重がんのナンバリングについては、やっかいな問題が発生する。一患者に1つの登録番号を与え、さらに腫瘍番号を示す接尾辞を付け加えることが、よく主張される。これは多重がんの間のリンケイジを容易にし、フォローアップを促進する。もう1つの方法は、多重腫瘍を有する患者に複数の登録番号を発行し、各々の原発腫瘍の中で相互参照データを供給することであるが、この手法は勧められない。

オンラインシステムの登録で有用な、もう1つの方法は、各新規のがん患者に、一連の登録番号を使って患者番号を与えることである。各腫瘍には1つの腫瘍番号が与えられる。この場合、第1番目の腫瘍には患者番号と同一の番号が与えられる。もしもその患者に別の原発がんが登録される場合には、同じ患者番号が使われるが、新しい腫瘍番号が割り当てられる（登録番号の次の番号）。データは、腫瘍番号を使って保存、処理される。しかし、一人の患者の種々の原発腫瘍は、同じ患者番号を持っているので、一緒にリンクすることができる。

情報保護

がん登録は、今日、個人データの情報保護に対して増大しつつある懸念を背中に背負いつ

つ、実施されている。全てのがん登録にとって、それぞれの患者を同定するのに十分な詳細情報を得ることが絶対に不可欠であり、それらがなければ、死亡診断書を通じて入ってくるものを含め、重複届出をリンクすることが不可能である。登録室の大多数にとって、これは各症例の氏名と、恐らくは住所を持つことを意味する。患者を識別する能力を持たずして、がん登録は運営できない。情報保護の問題については第15章で詳しく述べられる。

データの組織化

各々の登録原発腫瘍について、別々のレコードが作成される。それ故、多重腫瘍を持つ患者は、複数腫瘍のレコードを持つことになる。多重腫瘍の存在を示すのに特別のコードを使うことが推奨される（第6章参照）。腫瘍レコードに、コード化して収納する項目は、第6章で詳述されている。

データを組織化する方法は、データを機械的処理のためにパンチカードで持つか、それとも電算ファイルで持つかによって、大部分決定される。データの組織化の目的は、データの保存と抽出、及び解析を促進することにある。

データのコード化

データの組織化とは、通常、ある種のコード化のプロセスを意味しており、登録が手作業によるものであれ、電算機によるものであれ、基本的原理は同じである。

出来る限り、登録室は国際的に認められたコード体系を使用するように努力すべきである。第1に、これらの体系は、通常専門委員会で設計されており、その共同の知恵は、単一の登録室で得られるものをはるかに越えている。その結果、よりよい体系となっている。次に、国際基準を遵守することは、国際的な比較性を獲得する唯一確かな方法であり、主要データ項目について、国際的に合意のあるコード化システムを採用することは、登録室間の比較を行う上で、自明の利点である。推奨される種々の項目についてのコードは、第6章、第7章に記載されている。

登録室は、現場での限られたデータ項目、例えば病院や専門医をコードするのに登録室独自のコード体系を作成する必要に迫られる。1つのデータ項目で使うコード体系に、ある種の構造、もし可能なら、分類の要素を有するものを組み込むことは、よい考えである。後にデータ項目に関して要求される解析や選択の形式をできる限り考慮しておくべきである。コード体系をデザインする者の中には、コード化されるべき項目のリストをコンパイルし、項目をアルファベット順に並びかえてから、一連の数字コードを適用する傾向がある。

とりわけ電算機が使われている場合には、単に区別するだけの機能を越えて、コード体系を拡張することは、大きな価値がある。恐らく、1つの登録室に報告されてくる患者を治療

している全ての専門医を識別するのに、3桁もあれば充分であるが、さらにそのコードに4桁目を加えて、その専門医の専門領域、例えば、一般外科医、放射線治療医、婦人科医、など、を示すことは、価値がある。専門科別のがん紹介数を示す表の作成は、このコード体系が採用されておれば、非常に単純であろう。一方、それがなければ解析は、専門科を特定するために、非常に骨が折れるであろう。

コード体系をデザインするとき、コード化されるデータ項目を吟味して、その性質を理解することが大切である。1つの軸、言い換えれば1ディメンジョンで、全ての変数が分類できて、コード化も可能と考えてはならない。データ項目によっては、数ディメンジョンを有するものがある。例えば、診断、これは、腫瘍学のための国際疾病分類（ICD-0）では基本的に3つのディメンジョン、すなわち、部位、組織型、性状からなる変数であると認識されている。これらは、あたかも3つの独立変数であるかのように扱われ、それ故、どのような組み合わせでもコード化できる（第7章参照）。多次元データ項目の、もう1つの例は職業である。多くの職業は1つの産業でのみ従事されるが、一方疫学的研究では、例えば行程作業員が雇用されている特定の産業を知ることが、重要なことがある。コード化のレベルでこれに対処する最も満足すべき方法は、職業と産業を1つの2次元変数とみなし、そのようにコード体系をデザインすることである。

データ確認

データの質ができるだけ高く保たれるようにする事は、非常に大切である。この点は第9章でさらに詳しく述べられるが、うまくデザインされたシステムにおいては、ことに電算化されたシステムでは、データ確認は、データ組織化機能の一部である。定義により、不当なデータは組織化され得ないが、不正確なデータは組織化され得る。しかし、不正確なデータをすべて検出することは不可能である。すなわち、例えば、ある患者が、実際には1932年に生まれたのに、年の数字が置き換えられて、1923年6月15日生まれと登録室に報告されることがある。そのデータ項目は不正確であるが、正当であり、従って、もしも年齢が記録されていてクロスチェックに使われるか、もう1つのレポートが届いて、それが正しい誕生日を有しているのでなければ、誤りは恐らく気づかれない。日の数字の1923年6月51日への置き換わりは、不正確かつ不当であり、従って、決してデータベースに蓄えられない。システムは、不当なコードを含む不当なデータを、できるだけ早く検出するように、デザインされるべきであり、これは、登録のデータ組織化手法に組み入れられるべきである。

データ組織化の記録

登録が進展するにつれて、データ構造の変更が必要となってくる。新しいデータ項目が導入されて、新しいコードを作成することが必要になる。データユーザーが、データから何を期待すべきかが判るように、これらの変更の全てを完全に記録しておかなくてはならない。

不運なことに、何らかの新しい状況に反応して、あるいは個別のケースについての独断的な決定の結果として、殆んど衝動的に、変更が加わる場合がある。しばしば指示は口頭で、もしくは1枚の紙面に記載したメモの形で与えられる。いずれにしても登録室は、新規コードが導入され、旧コードを使わなくなった時、その日付を含め、データベース構造へのあらゆる変更の詳細な正式記録をもつようにするべきである。

コード化の大規模改訂

時代の経過と共に、コード化体系の中には大がかりな改訂を必要とするものがでてくる。登録室は、現場の体系の中で、これを避けることができるかも知れないが、一方、国際分類は、周期的に改訂され、登録室は従うように強制される。データの連続性を維持できるよう、旧レコードが新規コードを持つ様に変換されるかどうか、もしくはある時点で、できればある罹患年で、明確な中断が為されなければならないのか、2つ（もしくはそれ以上）の一貫した体系を使い得るのかどうか、熟慮されるべきである。

後者の方法は、例えば、データが手作業で処理されているか、もしくはコード体系が、意味のある正確な変換を可能にしない故に、データ変換が不可能な場合に限り、用いられるべきである。前期に属する症例に関する新規登録が、後期に属する症例と共に入ってくる時、ほぼ確実に、両体系は、ある期間一緒に使われるのであるから、不連続的なコード体系は、コード化の誤りの主要な源となる。そうしたデータの解析に当たって、またデータを維持するための電算システムのデザインにおいても、大変深刻な問題が発生してくる。新しいコード体系が考慮される場合には、それが従来の体系と、将来にわたって整合性があるのかどうか、確認のためのあらゆる努力を払うべきである。データ変換は、新規コード体系導入の見積りと計画をする際に、考慮すべき要因の一つと認識されるべきである。もしコード変換が実行される場合には、データの解釈に結果的に影響することが殆ど避けられないので、このことを完璧に記録しておかなければならない。

手作業で処理されるデータの物理的組織化

多くの登録室は、機械ソーターやタブレターで数えられるパンチカードの形で、データを維持しているが、電算処理の導入は、この器械類のほとんどを、時代遅れとしており、これらの方法をなお使っている登録室は、できるだけ早く電算化されるように強くアドバイスされる。

登録室によっては、エッジ・パンチカードデータが使われてきた。しかし、これらが現実的な将来性を持っているかどうかは疑わしい。それらは、たぶん年間1,000以下の小さな患者数についてのみ使い得る。資金が極めて限られているのであれば、1,000米ドル以下の費用の家庭用電算機を使うことによって、この程度の量のデータを維持することが可能であろう。

データが完全に手作業によって維持される場合、物理的に3つのファイルを持つのが伝統

的なやり方である。即ち、本章のはじめに記述したように、アルファベット順に整理した患者索引ファイル、登録名簿、及び固有のデータカードあるいは腫瘍記録である。

登録名簿は、単なる登録患者の一覧であり、登録順、すなわち登録番号それ自身により整頓される。この名簿は、実際、すべての新規患者に登録番号を与えるのに使われる。登録名簿には、少なくとも、登録年、登録番号、患者名、そして腫瘍の原発部位を、含めるべきである。

データカードは、各々登録される腫瘍の詳細を含む物理的記録である。これには、いくつかの物理的形態がある。登録抜粋様式の他、上述のように、種々のパンチカードが過去に使われてきた。通常これらの腫瘍記録は、部位毎に番号順に保管され、そのため肺、胃、等の記録カードのボックスが存在する。如何なる計算も殆んど部位カテゴリーによるので、これにより計算が容易になる。計算は、通常、性別にも行われるので、男と女とを別の色のカードで持つと、ある程度の利点がある。

電算化されたデータの物理的組織化

用い得る選択肢が多く、かつ各選択の影響が相当なものであるので、これは極めて複雑な課題である。この問題は、'Directory of Computer Systems Used in Cancer Registries (Menck & Parkin, 1986)' に、かなり詳しく述べられているので、ここでは概略のみ述べる。媒体の選択は、通常は、磁気テープかディスクかのいずれかである。

(1) 磁気テープによる保存

磁気テープファイルは一連のレコードからなり、各がん症例は、恐らく1レコードを占有する。一方、各磁気テープは何千ものレコードを含んでいる。ファイルは1リール以上に及ぶことがあるので、現実にはファイルの大きさのリミットはない。がん登録以外の応用では、数百万のレコードからなるファイルが稀ではない。時として、使用するプログラムが小さな制限を持つことはあるけれども、各レコードを80キャラクターに限定するパンチカードの古い制限は、普通、磁気テープのレコードには当てはまらない。磁気テープのレコードは順番に処理される。すなわち、レコードは磁気テープ上に保たれている順番に、読み書きされる。通常磁気テープ上のレコードは、元の位置では変わらない。もし変更が必要になったり、新しいデータが付け加えられる場合には、変更がなかった全レコードと共に、変更が加わった新しいレコードを含む全く新しいテープを書く必要がある。レコードは、単に古いテープから新しいテープにコピーしなかったことで欠損する。1つのテープから別のテープにデータをコピーするには、大きい電算機でも、1時間以上を要することがあるので、一瞬に1レコードを変更することは明らかに不可能である。それ故、変更はバッチで実行される。恐らく数千の変更が1行程で実行される。これはバッチ処理として知られており、磁気テープシステムの特徴である。保管の形態として、磁気テープは比較的安価であり、大抵の登録室は、

これを主要なデータ媒体として使用している。しかしながら、レコードは番号で並べられなければならないし、一方、大抵の分析はある特定の部位、もしくは部位群を指向しているので、とりわけ大きなテープファイルの分析を実行するには、便利ではない。登録室によっては、それ故、診断によって異なるファイルに編集された、例えば、1つは肺がん用に、今1つは胃がん用に、等と、手作業で維持するデータについて推奨されたのと殆ど同じ様な方法で、複製レコードを持っているところもある。

(2) ディスクによる保存

データをディスクに保存する為に使われる方法は、磁気テープに使われる方法よりかなり複雑である。磁気テープよりディスクの優れた点の1つは、ディスクの物理的位置にかかわらず、どのような順でもレコードを処理できることである。かくして、一瞬に1レコードを変更することが可能であり、通常オペレーターは、VDU を使って、データと直接通信することが可能である。1症例をスクリーンに表示させて、必要なら、システム中の他のレコードに干渉することなく、変更と書き換えができる。これはオンライン処理として知られている。そしてこれは電算機の効率的利用に関して、多くの新しい可能性を開いている。

Interactinal record linkage は既に議論されたが、ここでは強力なコード化技術を吟味する。データをどのような順でも処理できるので、たとえオペレータが知らなくとも、電算機はレコードが物理的にどこにあるのかを'知って'いなければならない。これはシステムが維持している索引ファイルの中で、ポインターを作ることで達成される。慎重にデザインされた索引化技術を用いれば、データはランダムに(オペレータがVDUを使って)、あるいは、例えば、番号、アルファベット、診断名、等、種々索引化された順で、アクセスされ得る。データは一回だけ保存され、使われる索引の各々は、全てのレコードのポインターを持っている。

この複雑性を有するデータベースを維持するために使われる多数の優れた商用のソフトウェアパッケージがあるが、一方で、これらを使うシステムの詳細な指定においては、相当な専門性が必要である。それ故、この種のソフトウェアのデザインに乗り出す前に、電算機の専門家のアドバイスを求めなければならない。

コード化の技術

(1) 手作業によるコード化

上述のように、コード化の主要目的は、分析を効率的にできるように、データを組織化することである。手作業によるコード化は、コード化マニュアル中にコードすべき用語を見つけだし、コードを記録することから成り、その点において分かりやすい。登録記録で使われている用語は、コード化マニュアルに必ずしも与えられているとは限らないので、実際は、経験と訓練、技量が要求される。従って、ICD-0 を使用するコード化事務員は、「Intra-

duct adenocarcinoma, invasive」と記載された腫瘍を、「8500/2, Intraductal adenocarcinoma」ではなく、「8500/3, Infiltrating duct carcinoma」とコードすることを、知っている必要がある。

誤り易く、かつ、多くは検出されない（後の段階で検出されるものもあるが）ので、コード化は、コード化事務員の側において、大変な集中力が要求される。疲労によって、受容できないほど高い誤り率をもたらされることがあるので、事務員が1日にコードする症例数の限界を設定しておくことが、多分賢明である。

がん登録業務の管理に従事する者は、コード化の正確性についてハイレベルの責任を担っている。十分なコードマニュアルが利用でき、これらが最新の、良好な状態に維持されていることが、絶対に必要である。適切な訓練が提供され、適切な監督が必要である。規則は十分に記録されなければならないし、大きな変更は、導入前に慎重に実施調査がなされなければならない。当初によく考えないと頻回に変更を加え、そのことが結果として、コード化職員にはいらだちとなり、いくつかの種類の誤りにつながり易くなる。

(2) 電算機によるコード化

オンライン処理の導入によって、手作業によるコード化の作業量を少なくしたり、全く無くしてしまった登録室がある。

電算機によるコード化の基本的理論は、手作業によるコード化のそれと全く同じである。即ち、辞書にある用語を検索して、適当なコードを抽出する。電算機を用いたコード化の際には、コードブックの内容がディスクに貯えられていて、オペレータは、普通はVDTを用いてコード化すべき内容を入力すれば、電算機が辞書にある用語を検索して、コードを与える。たいていのシステムでは「好ましい用語と同義語の方法」を用いている。使用される各々のコードは、1つの好ましい用語といろいろある同義語とに関連づけられている。これは例によって、よく示すことができる。

ICD-0 (WHO, 1976) の形態学の節の中で、8070/3 というコードは、「Squamous cell carcinoma, NOS」という好ましい用語と関連づけられている。しかし、他の用語も見出し語が下記のように与えられ、掲載されている。

8070/3 Squamous-cell carcinoma, NOS
epidermoid carcinoma, NOS
spinous-cell carcinoma
squamous carcinoma
squamous-cell epithelioma

2行目から5行目までの一段右によせた4行の用語は、すべて1行目の「Squamous-cell carcinoma, NOS」の同義語であり、すべて8070/3のコードが与えられている。それで、オ

オペレータが「Epidermoid carcinoma NOS」を入力すれば、電算機が 8070/3 のコードを付ける。このデータ項目が次に脱コード化されるときには、端末のディスプレイや解析結果の打ち出しとして、その「好ましい用語」である「Squamous-cell carcinoma, NOS」が表示され、元の用語は失われる。コード化マニュアルに掲載されている以外の用語は、承認された略字を含め、つけ加えることができる。例えば、「SCC」は殆んど間違いなく上の例では掲載されているであろう。いくつかの用語の最もよく起こる誤綴りと同じく、「NOS」を省略した形も同義語として載るかも知れない。電算化のコード化システムでは、コード化される前に、内容を編集するための処理手法をも含んでいることがある。これは、いろいろな状況で起こり得る省略形を、拡張するのに有用である。例えば、「Carcinoma」に対する「Ca」、あるいは、もしもこれらが辞書に掲載されていなかったならば、句読点もしくは「Gland」のような冗長な言葉を除くこと、など。

がん登録の記録に現れた用語が辞書に見いだせない場合には、当然、困難が持ち上がる。このことは、手作業によるコード化でも起こるが、この場合は、コードをつける人（コーダー）が、所定の用語に適用される最も適切なコードを選択する。一方、電算機によるコード化の場合には、所定の内容にふさわしい別の用語を選択し、入力しなければならない。これは、稀に、コードブックを参照することを意味するが、スクリーンに辞書の関連部分を表示する手法をシステムに組み込むことにより、スクリーンからオペレータが最も適切な用語を選択することができ、より便利になるであろう。新しい同義語は、コンスタントに辞書につけ加えられるので、「学ぶ」のは、オペレータというよりも、むしろ電算機である。しかし、オペレータの手腕の程度を低く評価してはならない。医学データのコード化は、しばしば必然的にある程度の解釈を含み、それ故これには、使われる用語の理解とその使用についての経験が必要である。電算によるコード化は、疑いもなくコード化事務員の効率を増し、コード化の正確性を確実に促進する。これは、職員の訓練が少なくすむとか、上級でない労働者でまかなえるということ、意味しているのではない。熟練の事務員が依然として求められるが、その人数は少なくすむ。

データ辞書

データ辞書の活用は、入力時のデータの組織化を、出力時のデータの組織化と関連づけることの重要性をはっきり示すので、そのことをここで考慮しておくことが適切である。

データ辞書とは、それぞれのデータ項目に対して、その名前、電算システムのどこにそれが貯えられているのか、入力時にどのように処理されるのか、出力時にどう処理されるのかを定義するテーブルである。また、データ項目に対して実行される Validity check（エラーチェック）に関する設定を含んでいることもある。データ辞書を使ってデータの組織化をする大きな利点は、プログラムの指示がアプリケーションと独立している点である。言い換え

ば、詳しい設定は辞書の中で定義されるので、多くのシステムを動かすのに1つのプログラムが使われるということである。これは、システムのハードコピー記録を提供するために出力され得る。もし変更が必要な場合、変更されるのはデータ辞書であり、実際のプログラミングは必要でない。

システムから情報を取り出すためには、情報がどのように格納されているのかを知っている必要がある。データ辞書は、そうした情報を提供する。解析ソフトは、ユーザーが、例えば、どの変数をクロス集計するのかを指定し、電算機が各レコード内の項目の位置を見出し、作表し、そして、結果を出力する時には、データ辞書中のコードに相当する用語をラベルとして使えるように、デザインすることが出来る。データ辞書は、例えば、項目の導入、使用中止、あるいはコード体系の変更時、など、システムに対する変更を記録する目的でも使い得る。

媒体変換

電算システムを使うときは、データを機械で読める形式に変換して、電算機に送り込む必要がある。この目的のためには、パンチカードがしばしば使われたが、今は、キーボードからテープへ、あるいはキーボードからディスクへの方式に変化した。

データを手作業でコード化するときには、コード化のための特別の用紙を使うか、元の記録用紙に組み込んだスペースにコードを書くか、いずれかである、コード化のための典型的な様式が、図3に示されている。誤ったレコードを修正しないように、各コーディング様式は同定番号を持つべきであり、また名前チェック（カラム 9-11）も必要である。ある文字のパンチに関して、例えば、ゼロとアルファベットのOとを区別するため、協定を採用することが大変重要である。曖昧な表現とパンチの誤りを避け、また、適切なパンチのスピードを維持するためにも、明瞭な記入が必須である。コーディングを抄録様式に行う場合、項目によっては、直接コードを選べるような様式にすることも可能である。例を2つあげると：

- | | | | |
|---|------|------|----------|
| 性 | 1. 男 | 婚姻状況 | 1. 独身/未婚 |
| | 2. 女 | | 2. 既婚 |
| | | | 3. 死別 |
| | | | 4. 離婚 |
| | | | 5. 別居 |
| | | | 9. 不明 |

その他の例は、第6章で示されたデータ項目に対して示されたコード体系から導き出される。これらの例では、コーダーが適当なカテゴリーをマークし、データ入力者が相当するコードを入力する。プレコードされた項目に、臨床医が直接コードをえらぶ方式は、コード化と転写の誤りを最少にするが、この方法は限られた項目にしか適用できない。より複雑な変数については、特別に設けたコードボックスにコードする必要がある。こうしたコードボッ

登録番号								名前			チェック	年齢	性	生年月日					
8	7	0	0	1	3	6	8	S	M	I	6	7	M	0	3	1	2	2	0
1							8	9	11		12		14	15					20
婚姻状態			診断日					部位			組織型								
M			1	3	0	4	8	7	1	6	2	9	8	0	7	0	3		
21			22					27	28				31	32					36

図3. コード用紙 (見本)

クスは、様式の辺縁か、コードされる内容の近傍に設けても良い。

コード化した様式は、キーオペレータにまわされ、キーオペレータは、文字あるいは数字データと共に、直接電算機にコードをタイプするか、もしくはパンチカードかテープあるいはディスク上に、レコードを作成する。後者は続いて電算機に入力される。パンチミス为了避免のために、それぞれの様式は、2度タイプするか、確認作業をいれる。そして、1回目と2回目の施行の相違が指し示され、どちらが正しいのかを判断するためにチェックされる。

オンラインシステムが使われ、データがVDUから直接電算機にキー入力される場所では、媒体変換は不必要である。修正が必要な場合には、その場で、その時に実施される。入力の際に、視覚によるチェックが行われるので、確認作業は通常は不必要である。

電算システムをデザインするときには、他の登録室あるいは研究機関にデータを転送するために、データを磁気媒体に出力する手技に、考慮が払われているべきである。比較可能なデータをプールする能力は、多くの研究において重要な要素であり、もしこれが磁気テープやフロッピーディスクを使うことで行えるのであれば、必要とされる仕事量は大幅に減少する。沢山の、同一ではあるが分離した解析結果を結合するよりも、単一の解析を実行する前にデータをプールの方が、一般的にいて、ずっと満足がいく。もしシステムが完全にディスクベースであれば、中間電算機を使う必要があろうが、データを磁気テープに抽出するためのルーチンは、電算化されたがん登録では、常にシステムの一部であるべきである、

問い合わせの発生と調査

問い合わせの発生

精度の高いデータを維持するためには、欠落情報の補足、もしくは既にあるデータ内に発生している不一致を解決する目的で、登録室は、しばしば追加の問い合わせをする必要がある（第5章参照）。手作業による登録では、これには、患者の情報と問題の性質を提示する標準の手紙を差し出す形をとる。高度に電算化された登録では、システム自身がデータの欠落あるいは不一致を検出し、自動的に問い合わせを発生させるルーチン業務を組み込んでいることもある。これらの問い合わせは、週単位に、あるいは月単位に打ち出される。図4に、未登録の患者について、死亡診断書を受け取ったときに、問い合わせを作成する方法の1例が示されている。一時的に死亡診断書で登録されるが、もし、がん患者が病院で死亡していたなら、死亡した病院へ、問い合わせの手紙が送られる（電算化されたシステムの場合には、自動的に作成される）。その症例は、病院では、通常の方法で要約が作成されている。もしがん患者が家庭か介護施設で亡くなったのなら、問い合わせは、通常は患者の家庭医である死亡診断書を書いた医師に送られ、患者が入院した病院の有無、病院名、などを、入院病院がない場合には、診断の基本的事項、とりわけ初診日を尋ねることになる。入院していた患者については、要約票を作成し得る関連の病院に向けて、さらに詳細な問い合わせが行われる。こうした作業は、手作業でも、電算機でも、なし得る。

情報の不一致が生じた場合には、これらは、通常は病院であるが、データ提供源で解決されるべきである。登録室では、正確な診断を決定しにくいことがある。このことは、再発か新規原発かの第2がんの判定の場合、症例の注釈に付けられた情報が、どちらとも言えない場合に、しばしば発生する。また、部位と組織型が一見不一致と考えられる場合に、電算機が登録を受け付けられない時にも、困難が生じる。こういう場合には、登録室の責任者が、患者を見ていた主治医に個人的に手紙を書いて問い合わせるという方針をとるのがよい。問題の良い解決法であるし、登録室の存在と登録室がデータをできるだけ正確なものとするように努めていることを、医師に知らせる機会にもなる。これらの問い合わせは、大量に生まれないうようにし、純粹に困難な場合にのみ行うことが重要である。このことは、持続的な関係を培い、信頼を発展させるという登録室の仕事の重要な要素を象徴している。こうした問い合わせにより、同じ症例もしくは似た症例に関して、その時には請求しなかった情報を、後に獲得できることがよくある。

フォローアップ（予後調査）

（1）能動的予後調査

能動的な予後調査システムを取っている登録室は、通常毎年、生存していると考えられる患者について、問い合わせをする。問い合わせは、通常初回治療の記念日毎に実施される。手作業によるシステムを組んでいるところでは、調査すべき全患者の索引カードが、調査予定の月毎のボックスに保管され、問い合わせの様式が、適切な時に、病院あるいは診療所に送られる。返送されたフォローアップ様式の結果、もしくは死亡診断書が届き、その患者が

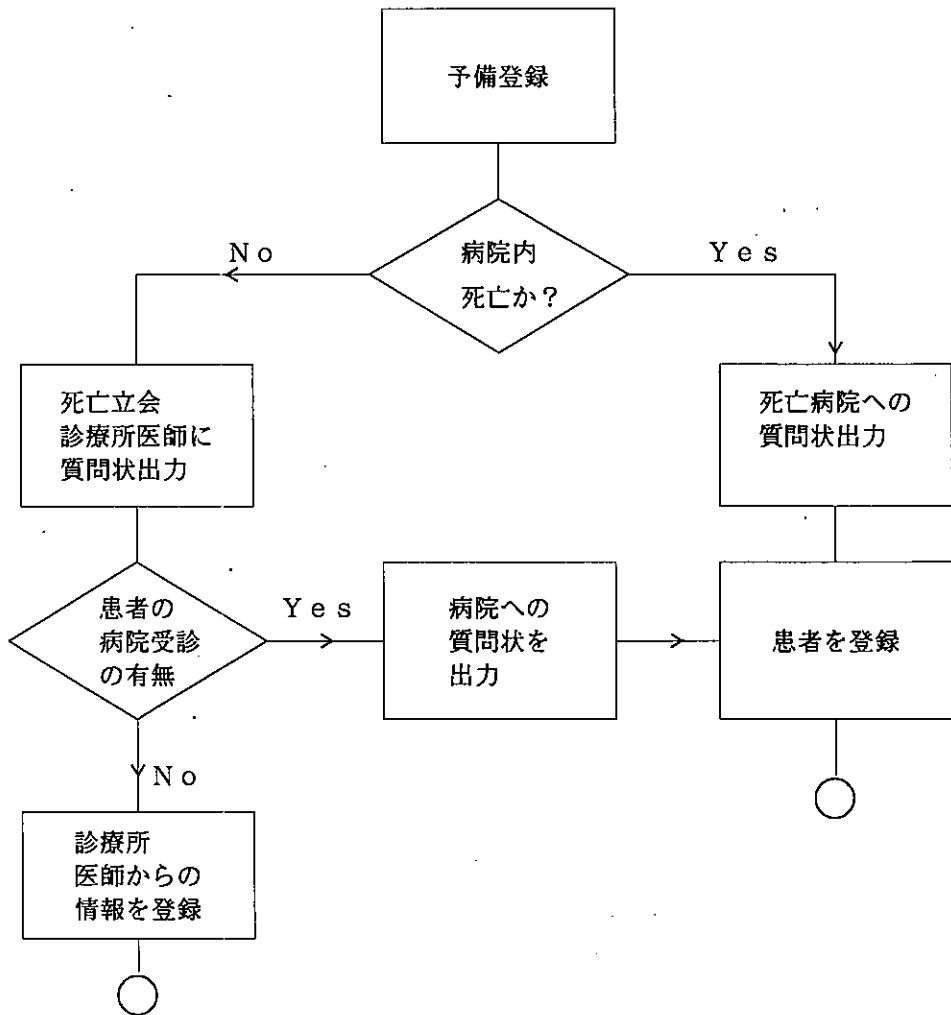


図4. 死亡票からの情報の取扱い

死亡と報告されたならば、患者カードがフォローアップ索引から除かれる。

バッチシステムを使う電算化の登録では、調査の要求が自動的に適当な月に、電算機から打ち出される。通常、これは登録室のアップデートシステムの中に組み込まれている。オンラインシステムをとっているところでは、記念日に基づいた索引ファイルを有しており、そこから、調査の要求が出力される。登録番号、患者氏名、住所、その他の必要なデータが、予め印刷された様式に打ち出される。これらの様式自身、手動でソートしなくてよいように、

住所地の順に出力される、こうしたシステムは、殆んど全自動である。そのため、これらを作成するのに要するスタッフの数は、少数ですむ。

(2) 受動的調査

受動の調査システムを取っている登録室は、がん死であれ、非がん死であれ、死亡場所がどこであれ、登録患者の全死亡の届出についての外部情報源に依存している。従って、ルーチンの問い合わせは発生しない。しかし、例えば、既に登録されている部位のがんと異なる病名が死因として上がってきたような場合には、特別の問い合わせが必要になる場合がある。

がん登録実施上の他の重要な側面

両者ともデータの物理的安全性に関連することであるが、あと2つの問題点を考慮しておく必要がある。

記録の管理

登録室に送られてくる全記録が、適切に扱われていることが大切である。大規模の登録では、存在する書類の量は非常に膨大で、情報が失われないことが、データの精度を高く維持する上で肝要である。情報が、登録室へ、もしくは登録室から送られてきた場合に、常に枚数を数えて、すばやく逸脱が発見できるようにする。処理済みの書類と、未処理の書類とは完全に別々にファイルしておくべきであり、このことは、適切な保存のための施設が必要であることを意味する。どの書類が、どの期間、どんな様式（マイクロフィルム化が必要になるかも知れない）で保管されるべきか、そしてどの書類が処理された後に安全に破棄されるのか、方針が明確に決定されるべきである。患者名のある書類を破壊する場合には、全て、安全で機密が保たれるよう、配慮されなければならない。情報を、適切な他の登録室に送る場合には、適切な方法が採用されるべきである。

書類及び電算化したデータの物理的安全性の確保

紙の書類及び電算ファイルの両方の紛失に対して、出来る限りの保護を実施しなければならない。情報保護及びデータそれ自身の価値の故に、このことは大切である。装置や建物は、紛失や損傷に対して保険を掛けることができ、これらを取り替えることが可能である。書類や電算ファイルの取り替えは、通常、不可能に近いので、紛失の機会が最小になるように警戒しなければならない。紙の書類は、コピーを維持しておくことは通常実際的ではない。火災、洪水、盗難に対して守られる条件下で保管されて初めて安全といえる。電算ファイルは、プログラムもデータについても少なくとも3つのコピーを持つべきである。ディスクのデータは定期的にバックアップを取って、ハードウェアの故障や事故による欠損の事態には、デ

ータが復元されるようにする。

データのコピーを、少なくとも1つ、登録室とは別の場所に安全に保管するべきである。電算化したデータについて、大規模な再構築が必要になった時には、元データの十分な数のコピーを作成し、何かうまくいかなかったときに、元データが復元されるようにしておく。データ変換の、わかりにくい重要な間違いが、かなり後になって解ってくることがあるから、これらのコピーは、永久に保管されるべきである。

電算化した登録室は、単にシステムを最新のものとする役目としてだけではなく、独立したプログラムとして、監査手続きを持つべきである。監査手続きを規則的に走らせて、データが不用意に消失していないかどうかを確かめるべきである。このことは、時に検出されないままケースが失われるマルチリールファイルを使う磁気テープシステムでは、特に大切である。遠隔地にある大型電算設備の比較的マイナーなユーザーである登録は、事故によるデータ損失に、特に脆い。消失するか損なわれる電算ファイルというのは、唯一、コピーがないファイルであるというのは、自然の法則と思われる！

(翻訳：津熊秀明、藤田 学)

第9章 質的精度とその精度管理

R. G. Skeet

他の何ものにも増して、がん登録室は、豊かな情報の源泉である。しかし、信頼性のない情報は、全く情報が無いよりも悪いと言われているから、どの登録室も、高い精度をもつ必要がある。データには必ず、質的な良悪があり、その良悪は、用いられた技術、方法でさまる。精度管理とは、それによってデータの質が測定できるシステムに対して与えられた名称である。登録室を運営するにあたり、精度管理のシステムなしに高精度のデータを作成することは、理論上は可能であるとはいえ、データが高精度であることを実証するために、精度管理システムの存在が不可欠である。大規模データベースは、完璧ではありえない。精度管理のシステムは、不完全な領域とその程度を確認するために設定されるが、それがデータの解釈をしやすくし、又、精度管理システムを変更する必要性を示唆することもある。

情報の質

情報の質は、データ自体の質とデータの発表方法の質との産物である。その考察にあたって、5主要領域に分類することができる。

患者把握の完全性

地域（一定人口集団）登録は、定められた人口中のすべてのがん患者を登録するべく努力する。この目的に向かって努力することは重要であるが、不注意によって生じる患者の重複登録を避けることも同様に重要であり、多くの登録室は重複が発生していないことを確認するための巧妙な技法を備えている。また別のエラーの源は、登録対象でない患者の混入にある。登録対象と定義されていない疾病の患者を誤って登録することもあり、その患者が登録対象地域内の居住者でないこともある。

詳細情報の完全性

全患者の全情報を常に確認することは不可能であり、全項目が全患者について得られるわけでもない。ある項目、例えば診断名や性別などは必須項目であるが、他の項目、例えば結婚歴などは必須でない、というように、システムをデザインすべきである（第6章参照）。

非必須項目に対しては、“記載なし”、“分類不能”、“不明”の別がわかるようにすべきである。無い筈のデータ項目に記入がある、という記載混入エラーも又存在する。これらのエラーは記載脱落のエラーよりも少ないが、それが発生すると、データの解釈が非常に困難になる。混入したエラー情報は、どこかから来たに違いなく、しかも同定されない誰かに

関係するデータではないかと思われる。

個々の情報の正確性

存在するデータすべてが正確である必要はない。個々の情報のエラーは次に述べる種々の時点で発生しうる；カルテからの抜粋、転記、コード化、及び穿孔の各経過でエラーがおこりうる。そのうちのあるものは、データの領域のチェック及び整合性のチェックで検出しようが、実際には不正確なデータでも、満足なデータであるように見える場合には、検出することは難しい。

報告の正確性

データベースが、多くの変数が入り、コーディングに不連続のものがあり、また、データが、異なるファイルレイアウトのままできも入っている、といった集合体である場合には、電算機から出る報告内容のリストとコード表などとの照合は、困難な作業となる。いくつかの登録室では、データのチェックをするプログラムが、電算ファイルの複雑さについての初歩的な知識もないスタッフの指示に従って、データの初歩的知識も無いスタッフによって作られている。このような状況下では、報告エラーも発生し易く、全く予期せぬ結果がもたらされない限り、エラーは検出されないまま、進んでしまうかも知れない。

解釈の正確性

登録室から発表される情報を適正に解釈するためには、データ源、データ収集方法および集計方法についての理解が必要である。このような知識は、経験と登録室のあらゆるレベルの活動に関与することによってのみ、得ることができる。またデータの正確性についての知識をも必要とするが、これは質的精度管理に関与することからえることができる。

質的精度管理

質的精度管理には、登録室の標準的な作業過程の一部を形成する日常プログラムとなっている場合と、データの質について特定の疑問が生じた時に処理する臨時調査とがある。あまり公式なものではないが、疑いもなく有用な精度管理は、データが使用される時に、これを用意深く吟味することである。実際、データを吟味しながら使用することが、精度管理の最上の方法の一つである、と考えられる。

患者把握の完全性の評価

患者把握の完全性（届出の精度）は、時々測定するというものでなく、常時、監視すべきである。これを行う方法の一つとして、受入れた死亡票について、未だ登録されていない患

者の割合を監視する方法がある。致命率の高いがんでは、この割合は相当高くなるが、より生存期間の長いがんでは、割合は小さくなる筈である。過去の実績から有意に逸脱しているようなことがあれば、登録室は、問題が存在する可能性があると考えらるべきである。これらは速やかな改善のための行動を必要とするかも知れない。でなければ、永久に登録されずに終わってしまうかも知れないのである。

最近の罹患数データを前年のものと比較することも、有用である。がん罹患率は比較的ゆっくり変化するので、何か著明な変化があれば、直ちにそれを究明すべきである。可能ならば、このタイプの監視は、部位別に行うのがよい。全登録数では十分安定している時にも、がん発生が稀な部位の一つに、登録数の突然の減少が発生する時があるが、これが検出されずに終るかも知れない。登録率の低下は、しばしば、ある部位に限って生じる。これは、例えば、ある研究者が特定のがんの研究を行っているために、カルテが定位置に返されていないからかも知れない。

これらのチェックの多くは、登録室の電算システムに組み入れることができる。というのも、このようなチェックは容易に自動化でき、特にこれについて考えをめぐらしている人がいなくても、定期的の実施しうるからである。

完全性の客観的な測定方法

登録の精度を計測するための方法は、様々提案されている。最も一般的なものが、死亡票を用いる方法 (Muir et al. 1987; Freedman, 1978; Benn et al., 1982)、又は病院カルテからサンプルをとる方法 (Chiazze, 1966)、である。これらの方法にも、明らかに限界があるが、登録室は、常時その精度を測定することが重要である。登録室が地理的に大地域をカバーしている時には、異なる病院からの通例の届出は、真の罹患数の変動の範囲さえも越えて、相当様々に変化する。この時には可能ならば、登録室の地理的領域の分割域ごとに、罹患率を定期的に計算すれば、可能な限り速かに登録率低下の可能性のある地域を同定し、改善対策をとることができる。

精度の程度はまた、しばしば診断に起因しがちである。悪性の記載のある全死亡票をルチンに受け入れる登録室では、脾、肺のような最も致命率の高いがんに関しては、実質的に完全にカバーされていることが多いが、非メラノーマの皮膚がんや早期の子宮頸がんではより不完全である可能性がある。

個々のがんの登録の精度を監視する一つの方法は、それらががんのために専門医を訪れる患者をサンプルとして、登録室に彼らが登録されているかどうかをチェックする方法がある。確認率を推測することは正確にはなしえないが、全ての登録室は、常識的な知恵と敬虔な希望に頼るよりは、このような客観的な計測方法を用いる筈である。

個々の情報の完全性と正確性

多くの登録室は、入ってくる全届出票を到着時に直ちにチェックする方法を採用している。少なくとも、最も重要なデータ項目について、全てその記載があるかを確認する。元の病院カルテがた易く入手できる時は、全てのエラーは、このようにして修正される。このチェックはまた、抜粋の仕方が良くないことを警告する役目をも果たす。届出票にある個々の項目が全て記載されているか、正確であるか、を知るための最良の方法は、その症例についてカルテの再抜粋と再コーディングを行うことである。これは、元の登録内容を参照せず、盲検法で行われるべきである。

最初の登録内容と再度作成した登録内容とを比較する時には、全データ項目を別々にチェックし、各項目についてエラー発生率を計算する。普通、各項目について、エラーの範囲を定めておく必要がある。なぜなら、不正確さは、しばしば程度の問題であることが多いからである。例えば、腫瘍の部位をチェックする時、ICDの最初の3桁で発生するエラーと、4桁目のエラーとは、区別することが望ましい。

この種の質的精度管理では、常時実施、時々実施の別に拘らず、チェックすべき登録の標本では、より一般的な腫瘍に対して重みづけをするのもよい。この方法では、同様の腫瘍の大数を再抜粋することを避けることができるばかりでなく、全体のエラー率を推定しなければならぬ時に、サンプルに対し適当な重みづけを適用することによって、全登録例での結果を再編成することが可能である。

Polissarら(1984)は、念入りの再コーディングの演習について述べて、結果の解析が複雑になりうることを示している。というのは、コーディングの不一致は、データ項目によってさまざまであり、これを制御するために標準化が必要になるかも知れない程である。

継続的、または臨時に行う精度管理

理想的には精度管理プログラムは各登録室のシステムの中に築くべきであり、それにより登録患者の一定割合のものが再抜粋され、再コード化されることになる。重大な項目、例えば診断名については、ダブルコーディングが全例について行われてもよい。これにより、コーダー間の一致度をも確認しうる。このようにして、データの精度の監視は継続的に実施され、日常の作業過程でのどのようなエラーも、速やかに修正される。継続的な精度管理計画を実施するとき、とりわけ、精度管理の仕事を1人の人にまかすのではなく、多数の経験あるスタッフが順番に分担する時には、高い精度を維持することの必要性を各スタッフに認識させることになる。唯一の短所は、経費のかかることである。もし登録室が、必要なスタッフ数をみだしていない異常な状態にあるのでない限り、追加の収入を見つけ出すべきであるが、その時には、チェックを常時行う計画よりも、臨時実施の計画の方が費用は得易いかも知れない。

臨時、又は継続的、いずれの精度管理の方法でも、エラーの程度を量的に測定するのみならず、データの正確さのレベルが常に改善されてゆくようなフィードバックメカニズムを取り入れるべきである。精度管理では、どの特定のデータ項目が、しばしば未記入のままであるか、或は著名に高いエラー率を示しているか、を明かにすべきである。一連のデータから、その項目を除いてしまうという提言が、考慮されねばならない。多くの登録室が、説明のできないようなデータ項目を収集しつづけていることは疑いの余地がない。これらの項目が収集対象から除外されるなら、有意義な経済的な節約となるだろう。

データ精度管理のための電算機チェック

がん登録室が電算化されているところでは、重要な2つのタイプのチェックを組み込むことができる。；正当性のチェックと一貫性のチェックである。

(1) 正当性のチェック

各データ項目に関して、正当でないコードがデータベースに供給されていないかどうかを確認することは、電算機によって実行できる。領域チェックの形をとってもよい。例えば、患者の年齢は0より小さい筈はなく、また、通常105より大きくはない。データ項目のフォーマットも点検できる。例えば、患者名はアルファベット文字のみであり、年齢は数字のみであることを確認する、など。

電算化された全ての登録室は、コーディング制御ファイルを持つべきである。即ち、電算ファイルに各項目についての正しいコードを入力しておく。新規に入力されてくる各コードは、この管理ファイルと照合され、適当でないコードは拒否され、報告される。

(2) 一貫性のチェック

これらのチェックでは、ある項目の内容と他の項目の内容とを比較する。よくわかる例をあげると、陰のう腫瘍が女性に、又は卵巣がんが男性に記録されていないことをチェックする。日付の順番では、誕生日、診断日、治療日、そして死亡日の順が守られていることを確かめるためにチェックされるべきである。腫瘍は誕生日時に診断されることもありえ、死亡後に診断されることもありうる。しかし、数日以上おくれることはないことにも注意しておく。収集するデータ項目が多い程、もとより、より多種類のチェックが可能となる。

ある場合には、エラーの可能性があるとするサインや警告が出て、これに注意がひかれるだろう。男性の乳がん症例や、小児での癌発生に警告が出されるかも知れない。それらは必ずしも誤りではないが、手作業による確認が必要なほど、稀なものである。一貫性のチェックの例を付録3(b) (テムズ登録室) の章に示し、テムズがん登録室の電算システムによって作られたエラーメッセージを表1に示した。

電算機によるデータチェックは非常に効果があり、オンライン(データ入力時)、オフライン(バッチ処理の一部として)、何れの場合にも実施しうる。後者では、データの修正は、

表1. テムズ地域がん登録室（英国）のコンピュータシステムにおけるエラーおよび注意メッセージ（論理チェック項目）

1. 生存者の死亡情報	24. 同定情報なし
2. 死亡者が死亡情報をもたない	25. 病院欄に手術した病院名がない者
3. 単一項目に2種の情報	26. 病院欄に放射線治療病院の名前がない者
4. 診断日以後の誕生日	27. " にアイソトープ治療病院名がない者
5. 予後情報のないもの	28. " " に化学療法病院名がない者
6. 最終レポートの日付が、誕生日より前	29. " " に死亡病院名がない者
7. 最終レポートの日付が、診断日より前	30. 年齢と生年月日との不一致
8. 病院受診日が誕生日より前	31. 計算した年令
9. " が死亡日より後	32. 原発部位と性との不整合
10. 手術日が誕生日より前	33. 原発部位と同じ部位での多重がん
11. 手術日が最終レポートより後	34. 多重がんでの性と部位との不整合
12. 放射線治療日が誕生日より前	35. 部位の記載なし
13. " が最終レポート日付より後	36. 職業従事日の記載が正しくない
14. アイソトープ治療日が誕生日より前	37. "学生" は職業ではない
15. " が最終レポート日付より後	38. 16才以下で独身でない
16. 化学療法治療日が誕生日より前	39. 記載された日付に再度注意
17. " が最終レポート日付より後	40. 採録員のデータが正しくない
18. 治療なしとした日が誕生日より前	41. 職業の記載なし
19. " が最終レポート日付より後	42. 16才以上で、職業欄が「小児」の者
20. 臨床情報なし	43. 名前と性の不一致
21. 臨床記録の中の診断日と同定欄のそれとが不一致	44. ICD-Oコード195の者
22. 死亡していない人の剖検所見	45. 199.9とコードされているリンパ腫
23. 剖検診断日が最終レポート日ではなく、診断日の者	46. 部位と組織型の不一致
	47. 不適当な部位の良性腫瘍
	48. 男の主婦
	49. 多重がんの誤ったコーディングがないか

次のバッチ処理の機会に行われる。システム設計により、何らかのエラーを他のエラーより重大なものとして認識させることもできる。そしてエラーの程度を反映する何らかの尺度がセットされ、主要なエラーがあれば、その症例の登録そのものを拒否するが、より小さなエラーではデータベースにそのデータの修正を許す。このようなデータには、エラー項目を持っていることを示す旗をたてるべきである。むろん、最も重大なエラーから順に修正するように、優先順位を与える。

精度管理のための必要条件

(1) ルールと文書化

あるデータ項目についての2つの意見のどちらが正しいかを定めることは、もし、しっかりしたルールがなければ、不可能である。それに基づきデータが収集されるルールには、全項目にわたる学術的な定義と、その関連用語とを含めねばならない。ある症例には、主観的な判定が必要な時もあるだろう。これらは、常にスタッフの上級職員と相談しつつ、なされねばならない。決定の理由は文書化しておくべきで、これにより、将来同様な状況が発生した場合にも、同様に対処することができる。

(2) 良いコーディングシステム

良いコーディングシステムでは、どんな用語もたゞ1つのコードに割り当てられる。全ての用語を明確にコード化することが可能でなければならない。特に、明確な「否」がない時には「記載なし」が「諾」を意味することになる、あるいはその逆が成り立つような環境では、「記載なし」と「不明」の意味に特に注意する必要がある。

コーディングシステムが時期により変化している場合には、一定のコードが使われていた各時期に関し、文書化されたルールを持つことが必須である。例えば登録室が外科的手術のコードを変更した場合、コード変更はコーディングが行われた時期によるのか、手術の時期によるのか、その症例が最初に登録された時期によるのか、が明らかでなければならない。

(3) 基準

登録室が作業の基準を持つことが重要である。最大許容エラー率を、主要データ項目に対し定めておくべきである。例えば ICD-0 の3桁レベルでは 5%、或は性別では 0.5%など。これらの率を越えた時には、許容範囲までエラーを減少させる処置を、直ちにとるべきである。

その他の情報

「がん登録室の精度管理」(Statistical Analysis and Quality Control Center 1985) は、使用した基本原則と方法との分野に関してを記述した、精度管理への総合的なガイドブックである。精度管理に関する話題についての数多くの論文と、トレーニング用の演習課題の抜粋が含まれている。

(翻訳：花井 彩、藤田 学)

第 10 章 集計結果報告

O. M. Jensen and H. H. Storm

がん登録の主な目的は、特定の集団におけるがんの発生の統計を作成することである。発見結果と結論とを種々の形式の報告にまとめて、がん登録データの利用者へ配付しなければならないので、収集した情報の集計表作成、検定、および解釈は、がん登録室の活動の重要な部分を占める。データを使用して種々の形式の報告を提示することは、がん登録室の設置の正当性を示すために必須である。

がん登録室がもっている情報は、主としてがん罹患報告書、特定の主題に関する（特別）報告書、および科学雑誌掲載論文等によって伝達される。従って報告の形式は、単なる集計表から、例えばがんの発生や治療の結果についての仮説を立て、それを検証する高度な解析まで広がる。さらに、がん登録室のデータを報告することが、がん登録過程の質の改善に間接的に寄与する。なぜなら、データが集計されて、はじめて、がん登録室の入力運営上のエラーと矛盾を発見することは、よく経験することである。この章では、がん登録室からよく出される種類の報告、特にデータの集計表およびグラフによる提示について述べる。

がん罹患報告書

がん罹患報告書は、がん登録のデータの基本的な提示方法である。がんを届出してくれた医師、保健機関、および社会へ、がんの発生状況に関する情報を還元するものである。従って、がん罹患報告は、国や地域の保健情報システムの一部として重要な機能を果たす。さらに、がん罹患報告書に含まれている集計表は、そのがん登録が出すその他の報告の基礎データとなる。

がん罹患報告書の内容を決定する前に、報告書を毎年作成するのか、数年間分の罹患情報を基礎とするのか、を考慮することが重要である。毎年データを報告することにより、連続した還元が行えるが、多くのがん登録における多くの部位のがんの発生は稀であって、毎年の発生数が偶然変動の影響を非常に受けやすいことを認識しなければならない。従って、その地域集団での観察年数とがんの発生数とに応じて、例えば3年間、または5年間に、観察期間をまとめて報告するか、あるいは部位をより大きなカテゴリーにまとめてデータを提示する方が望ましい。もしくは、毎年の報告を、5年毎の詳細な報告で補足するという方法もある。これにより、特定部位についての成績をも含めた、より安定した結果を提供できる。

どの情報を伝達するかということと、この目的を達成するための最も適した形式、とを決定することが、非常に重要である。表の種類と構成は、その後にデザインできる。グラフを用いることは多彩になるだけでなく、表を読むことが難しい人に、非常に役に立つところが

大きい。提示の形式を決定するにあたっては、比較性のがん統計における重要な点であり、がん登録は長期にわたって運営されるものであることを忘れてはならない。従って、データ提示の形式は、長期間維持されなければならない、他の登録からの報告と容易に比較できるように、十分詳細に提供しなければならない。形式を変更しなければならない場合は、以前の報告に載せられた数を読者が変換できる形で、情報を提供すべきである。

報告が毎年か、例えば5年毎かによって、内容の細かさは異なるが、がん罹患報告書は、以下の情報を含むべきである。

- (a) 背景情報
- (b) 結果の提示と評価
- (c) 付表

読者が結果を解釈し、他の登録と比較するのに役立つように、報告書には背景情報を含めるべきである。データは報告の付表の部分で提示されるべきである。最後に、集計からの重要なメッセージを際立たせるグラフも含んだほうがよい。

背景情報

(1) がん登録室と登録手順の記載

がん登録室の機構の概要は、少なくとも2～3年毎に掲載し、掲載しない年には参照すべき掲載年を記載すべきである。登録室の専門職員と、それぞれの専攻、又は責任分野、例えば疫学者、統計学者、腫瘍学者等、のリストを記載すべきである。

登録手順の記述には、登録されている症例の情報源に関する情報や、実施している情報収集方法を含めなければならない(第5章参照)。登録対象となる疾患のリストも提示されるべきであるが、例えば国際疾病分類(ICD-9)(WHO, 1977)等を参照と略することもできる。登録とコーディング手順の簡単な説明は、読者が提示されている資料の質を評価するのに役立つだろう。

届出される疾患の種類が、登録によって異なるので、報告書に含まれているがんの明確な定義も与えられるべきである。がんの分類にICD-0(WHO, 1976b)の部位分類と形態学分類とを用いていることが多いかも知れないが、がんの定義はICD-9(WHO, 1977; 第7章参照)の140-209と230-239に限定されるべきである。もしがん登録室がICDと異なる分類を使用している場合は、2～3年毎に分類表を含めるべきである。がん登録は、部位によっては、がんか非がんかが論争されている腫瘍についての届出を受け取ることがある。報告書には、そのような腫瘍が集計に含まれているか否かを明示すべきである。例えば、尿管の良性乳頭腫(グレード0の移行上皮がんともいう)(ICD-0:M8210/1)と膀胱の浸潤がんを区別するのは難しい。WHOは全ての膀胱腫瘍はまとめて扱うことを推奨している(Mostofi et al.,

1973)。一般的には、もし必要ならば、読者がそのような論争的になる腫瘍を集計から除くことができる形で、データを作表すべきである。

一般的に認められている規則がない場合には、特に、報告書で用いている定義を明示しなければならない。例えば、死亡診断書のみからの発見例や偶然発見例（解剖や検診等）も、集計表の中に含まれているのか、細胞診は顕微鏡下での確診に含まれているのか、脳の良性腫瘍や性状不明の腫瘍は、悪性と診断された腫瘍と合わせて報告されているのか、膀胱腫瘍は乳頭腫を含んでいるのか、等は明示すべきである。多重がんの定義と取り扱いについても、罹患報告書に記載すべきである。

前がんやがんの疑いと記載されている記録を受け取り、保管している登録室も多い。そのような症例はICDの悪性腫瘍の範疇から外れているので、がん集計に含めるべきではない。それらの非がん状態が完全に登録されている場合は、罹患報告書の中で分けて集計することが可能である。

(2) 登録が対象としている地域の人口

罹患報告には、がん登録が対象としている地域の定義と、可能ならば概要を含むべきである。

罹患報告書の中でサブグループごと、例えば地区別や人種別等の罹患状況を報告している場合は、人口データの情報源も全部記載すべきである。市部/郡部比を提示している場合は、用いている市部と郡部の定義も特定しなければならない。

分母となる人口データの情報源は、参考資料も含めて記載しなければならない。付表には、罹患データの表に用いているのと同じ年齢区分別、サブグループ別の人口をも含めるべきである。付表にそのような表を載せるのに加え、背景情報の項で、人口ピラミッドを載せるのも役に立つだろう。

(3) 統計用語

罹患報告書作成用に用いられる統計手法も含め、がん登録室でよく用いられる統計手法については、第11章で詳細に述べる。どの罹患報告でも、統計用語と、年齢調整に用いた標準人口とについて記載した、簡単な章を含まなければならない。直接法では世界標準人口（第11章参照）が広く用いられている。この標準人口を使えば、異なるがん登録からの報告データを、読者が相互に比較することができる。

結果の評価

定期的な罹患報告の主な目的は、がん登録過程からの結果の伝達であり、読者が、それぞれの意義について結論を導きだせる形で、提示されなければならない。

情報は、報告に載っているデータを読者が利用するのに役立つように提供されるべきである。従って、がん登録室（関係者）には明白であっても、使用されているがん登録手法について詳しい知識がない読者には簡単に評価しにくい、結果や警告についても記述すべきである。

届出方法や登録手順について、診断の精度や届出の精度に影響を与えうる変更が少しでもあった時は、簡単にふれておくべきである。がん登録結果の報告に際しては、特に以下の点に注意を払うべきである。

(1) 毎年の症例数の定常性：新しい登録では、初めのうちは症例数が過大になることはよくあることであり、数が安定するまでは報告を差し控えるのが賢明である。しかしながら、運営の2年目、3年目に数が減少することもあり、これは、初期に、罹患者と同様に有病者が届出されていたことを示唆する。データの収集方法によっては、前年より症例数がやや少ないことを発見するがん登録室もあるかもしれない。差が大きい時は、まだ出版の時期ではないことを示している。突然、明らかに数が減少したときは、届出方法が崩壊したことを示しているかもしれない。発生が稀な部位では特に、数の変動がありうることに留意しなければならない。

(2) 部位分布：部位ごとの頻度が変わった場合（例えば、数が一致しない、あるいはある腫瘍が出現しなくなった等）には、必ず、その妥当性を認める前に、十分検討しなければならない。そのような状態は、さまざまな要因によって引き起こされる。例えばコーディングの誤りや、医師の関心が最近トピックになった腫瘍に向けられている場合、などが考えられる。

(3) 診断の精度の指標：2つの指標が一般的に用いられている。即ち、顕微鏡的に診断が確定された症例の割合と、死亡診断書のみによって登録された症例の割合である。がん登録の診断情報の精度の評価指標となるのに加えて、これらの指標は登録の完全性を評価するのに役立つ。即ち、顕微鏡的診断率が、全部位あわせて100%に近かったり、あるいは死亡診断書のみにより認知された症例の割合が大きい（例えば15%以上）場合は、報告もれがあることが考えられる。逆に、死亡診断書のみ割合が非常に低い場合（1%以下）（もし非常に効率的な遡及調査手順がなければ；第5章参照）、がんと診断された死亡診断書のうち登録室に届いていないものがあることを示唆する。

さらに、原発部位が不明と診断された症例の割合も、評価する価値がある。もし高ければ（独断的に10%とするが）、診断部門の能力が不十分であること、利用できる診断部門が十分利用されていないこと、あるいは結果の記録が不備であること、などのどれかを示唆しているかもしれない。

(4) 人口統計学的データ：性、年齢、あるいは居住地不詳の割合が多ければ、届出が不完全であることを意味し、登録室の職員からのデータ補充依頼が不完全であったことを意味

する。

(5) 類似した地域との差： 類似した地域から報告されている数値よりも、全がんの率がかなり低ければ、報告もれのあることが推察される。

表の作成と提示

がん罹患報告書の重要な部分は、表の部分である。表は、通常1つの章にまとめて、文章部分の後に提示する。

表の目的は、読者がその表から直接に、あるいは、さらに計算を加えて、結論を導き出すことができるように、結果を簡単な形で表現することである。作表は、コンピュータを利用するとかなり容易であるが、種々のパンチカードに情報を入力することによっても達成できる（一例が WHO Handbook for Standardized Cancer Registries (Hospital Based) (WHO, 1976a) に掲載されている）。

がん登録の結果の表の基本は、度数分布表、即ち、ある特徴を持つ個人の数の分布を示す表である。作表に関する基本原則は Bradford Hill (1971) が与えている。ガイドラインの要約と罹患報告の典型的な表の例を以下に示す。

- (1) 表全体の内容と各カラムの項目とを、明確に、十分に、定義すること。
- (2) もし率が表に含まれているならば、もとなる分母を明示すること。
- (3) 度数分布は完全に提示すること。
- (4) 率や割合を提示する場合は、必ず、そのもとになった観察数に関する情報を、何らかの形で提示すること。
- (5) 登録例を考慮の上除外した場合は、その細かい内容と削除の理由、削除基準を明示すること。

基本の度数分布表では、その期間の登録数を、がんの部位 (ICD)、年齢、性別に示す。例を表1に示す。年齢に関する情報は、5歳階級で表されるべきである。0-4歳のかわりに、0歳及び1-4歳を使用することもある。症例数が少ない場合は、10歳階級でもよい。この場合は WHO の推奨する年齢階級、即ち、0-4、5-14、15-24、25-34、等に従うべきである。解剖学的部位は、ICDの3桁分類に応じて提示されるべきである。表には、ICDによって組織学的に定義されたカテゴリー（第7章参照）も含むべきである。ICD-0の部位分類の軸 (axis) のみの表では不十分である。ICD分類と異なる場合は必ず、脚注を用いて明示すべきである。

表2のように、基本の度数分布表に付随して、同様の形式で年齢別、性別、部位別の1年間の罹患率を提示することもできる（率の計算方法は第11章参照）。年齢別の率は、数年間累積したデータのみを利用するのが好ましい。なぜなら、多くのカテゴリーのがんの1年間の発生数は非常に少ないので、計算するのに適当ではない。部位ごとの全年齢の粗罹患率とともに、年齢調整罹患率を提示すべきである。累積罹患率を表に含めることも、考慮すべき

表1. 1983-87年の間の新患数—デンマーク、男、部位別、年齢階級別

ICD-9 部位	0-4	5-9	10-14		80-84	85-89	90+	不詳	全年齢
140 口唇	・	・	・	(中略)	55	16	15	0	552
141 舌	・	・	・		9	3	1	0	177
142 唾液腺	・	・	0		14	7	0	0	112
143-5 口腔	・	0	・		16	12	7	0	350
146 中咽頭	・	・	・		12	2	1	0	218
(中略)									
206 単球性 白血病	2	・	0		4	1	・	0	24
207 その他の 明示白血病	0	・	1		9	5	5	0	90
208 細胞形態 不明白血病	1	・	1		9	5	3	0	63
195-9 その他、及び 不明確な部位	4	2	2		236	124	58	0	2119
・ 全部位	136	104	105		6077	2859	903	0	61711
・ 皮膚を除く 全部位	136	103	101		5235	2407	734	0	53169

訳者注：① 原本の表1, 2には、ICD-9の3桁分類で殆んどの部位が入っているが、上の表では、147~205の部位の成績を省略した。

② 15~79才の間も、5才間隔で区切られているが、省略した。

である。累積罹患率は、集団を比較するのに非常に便利な指標である。累積罹患率は生涯を通じての、そのがん罹患する期待値の概数であり、一般の読者に理解しやすい。

基本表は、人口のサブグループ別、例えば、市部及び郡部、地域別（地区、国、政令都市）等の同様の表で補足してもよい。分母となる人口も、同じ表中に載せるべきである。

罹患報告書における診断の精度は、診断の根拠を部位別に表にして掲載すべきである。表3に例を示すが、少なくとも、組織学的に診断された割合と死亡診断書のみの割合は、提示すべきである。

図の作成と提示

図の利点は、表よりも容易に注意を引きやすいことであり、傾向や比較を、より鮮明に示し、結果を思い出しやすくする。ひとつの図（グラフ）は千の言葉に匹敵する。統計表は、非常に圧縮された形式の中で多くの情報を提供することができ、情報を正確な値で詳細に提供できる。しかしながら、「どんなに明快に構成された表であっても、そのような提示方法は読者にいつも困難を強いる。」(Bradford Hill, 1971)。図は隠れた事実を表面に出し、解析的思考を刺激することができるが、いくつかの基本原則を忘れないことが重要である。

表2. 年齢階級別罹患率、粗および年齢調整ならびに累積罹患率—デンマーク、1983-87年、男、部位別、年齢階級別

ICD-9 部位	年齢階級別罹患率						全年齢		累積率	
	0-4	5-9	10-14	85-89	90+	不詳	粗率	年齢調整率 ^{a)}	0-64	0-74
140 口唇	・	・	(20.6	54.9	0	4.3	2.9	0.16	0.33
141 舌	・	・	中	3.9	3.7	0	1.4	1.0	0.08	0.12
142 唾液腺	・		略	9.0	0.0	0	0.9	0.6	0.04	0.07
143-5 口腔	・	0.0)	15.4	25.6	0	2.7	2.0	0.13	0.24
146 中咽頭	・	・		2.6	3.7	0	1.7	1.3	0.09	0.15
(中 略)										
206 単球性*	0.3	・	0.0	1.3	・	0	0.2	0.1	0.00	0.01
207 その他の*	0.0	・	0.1	6.4	18.3	0	0.7	0.5	0.02	0.05
208 細胞形態不明*	0.1	・	0.1	6.4	11.0	0	0.5	0.3	0.01	0.03
195-9 その他、及び不明確部位	0.6	0.2	0.2	159.5	212.1	0	16.6	10.5	0.49	1.18
・ 全部	19.4	12.8	11.3	3677.4	3302.2	0	484.0	312.0	5.52	6.20
・ 皮膚以外の全部	19.4	12.7	10.8	3096.0	2684.2	0	417.0	268.6	3.26	1.24

* 白血病

a) 人口10万人あたり訂正罹患率 (世界人口)

訳者 注: 表2の形式については

表1脚注を参照されたい。

- (1) ダイアグラムの唯一の目的は、目によって幾つかの数の意味を把握することを助けることである。従って、ひとつのグラフに示すデータ数は限定されるべきである。
- (2) 図は、必ず情報の補助とみなすべきであり、関連性や傾向の証拠となるものとみなしてはならない。それらの証拠は、主として統計表そのものから導きだされなければならない。従って、図単独の掲載は、受け入れ難い。つまり、図の基礎となった表の情報、必ず提供されなければならない。
- (3) スケールの選択により、同じ数字は違って見えることがある。
- (4) スケールの問題は、ひとつの図の中での比較でも重要となる。
- (5) 図は、それだけみれば内容がわかる形式、つまり図から得られる内容は、文章を参照しなくてもわかるような形であるべきである。

表3. 新患者の診断の根拠—デンマーク、1983-87年、男、部位別

ICD-9	部位	新患数	a) 組織診	b) 手術、 内視鏡	その他 不明	死亡票 のみの者
150	食道	737	93.8	0.3	1.9	2.2
151	胃	2569	90.6	0.4	3.1	3.0
153	結腸	4233	91.3	0.4	4.2	2.0
154	直腸	3458	95.0	0.2	1.9	1.2
155	肝	774	92.0	0.4	0.6	5.7
156	胆のう	377	84.1	0.3	6.1	2.4
157	膵	1842	76.0	1.1	9.3	3.8
161	喉頭	993	98.5	0.1	0.1	1.1
162	肺	11352	87.1	0.4	0.5	3.7
174	乳房	89	92.1	0.0	0.0	1.1
185	前立腺	6731	88.7	0.2	1.4	2.4
188	膀胱	4778	98.1	0.1	0.4	0.9
193	甲状腺	170	98.2	0.0	0.0	1.2
200.2	非ホジキン リンパ腫	1367	98.2	0.0	0.1	1.3
204	リンパ性白血病	881	94.0	0.0	0.1	4.8
205	骨髄性白血病	685	97.2	0.0	0.0	2.5
・	全部位	61714	91.4	0.3	1.3	2.3
・	皮膚以外 全部位	53171	90.2	0.3	1.5	2.7

a) 細胞診を含む。血液のがんの場合は、骨髄および末梢血の検査を含む。

b) 組織診なし。

訳者注) 診断の根拠とする検査は、表3の左から順に優先採用する。

よく用いられる図の例を以下に示す。図およびその作成のより詳細な記述については、読者は、例えば Bradford Hill (1971)を参照すること。

棒グラフ、もしくはヒストグラムは、数値データや順位データの頻度、割合、および百分率を表すのによく用いられる。棒は、縦も横もあり、その長さによって大きさを示す。図1に、がんの部位別の新患数の例を示す。年齢階級のような順位データは、その名が示すように、一定の順番で表されるべきである。

棒グラフは、1変数以上、例えば、部位—治療分布のように、変数ごとに違う色や網かけを用いて表すことができる。例を図2に示す。しかし、図を重ね合わせた形で表示しないことが重要である。

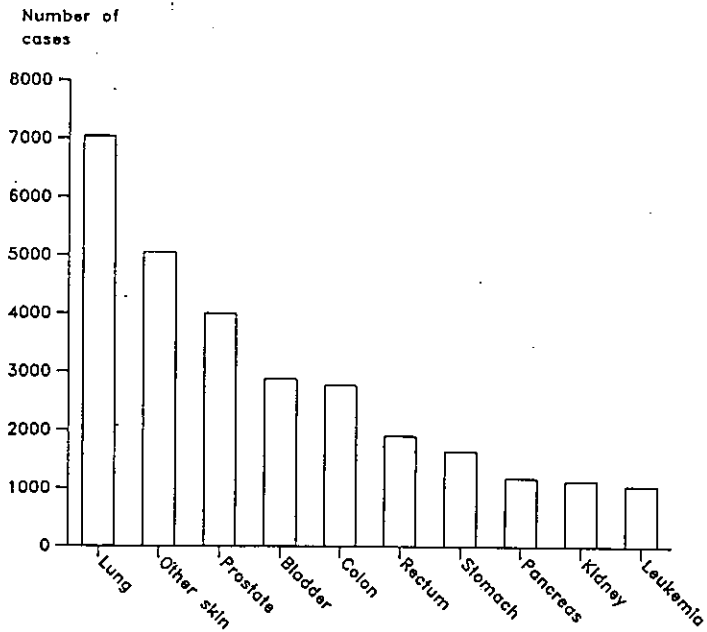


図1. がんの部位別罹患数—デンマーク、1983-85年、男、上位10位まで (棒グラフ)

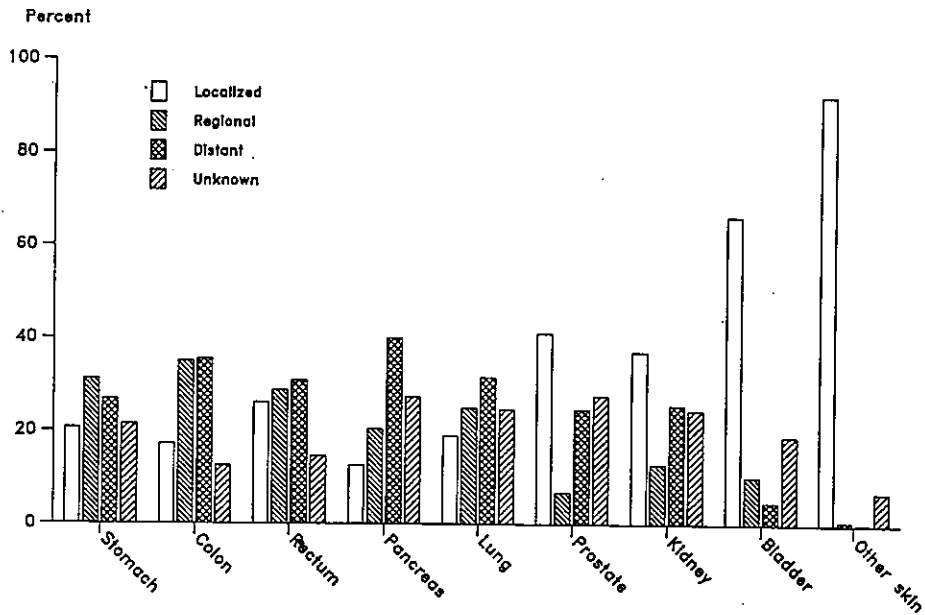


図2. がん患者の臨床進行度別の分布—デンマーク、1983-85年、特定部位 (棒グラフ)

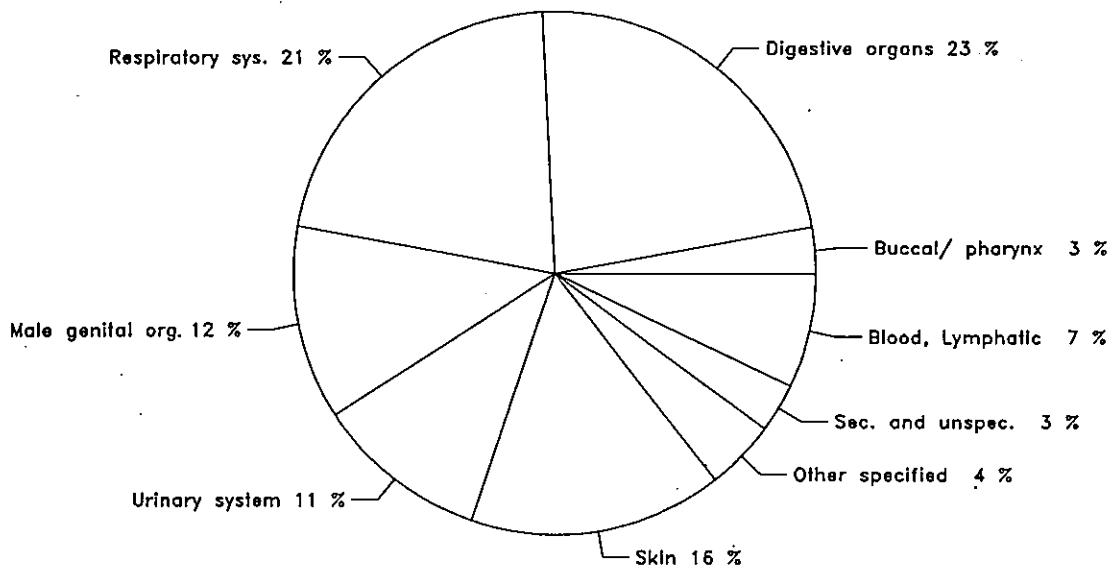


図3. がん罹患者の部位別割合 (%) -デンマーク、1983-85年、男
(円グラフ)

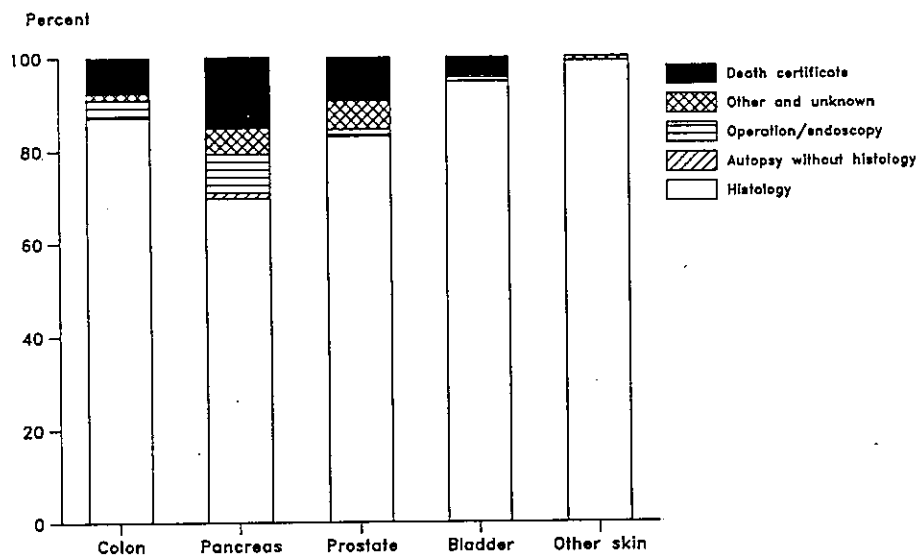


図4. 死亡票から把握した例について、死亡後遡及調査を行う前に判明していた診断根拠の百分率分布 -デンマーク、1983-85年、特定部位
(帯グラフ)

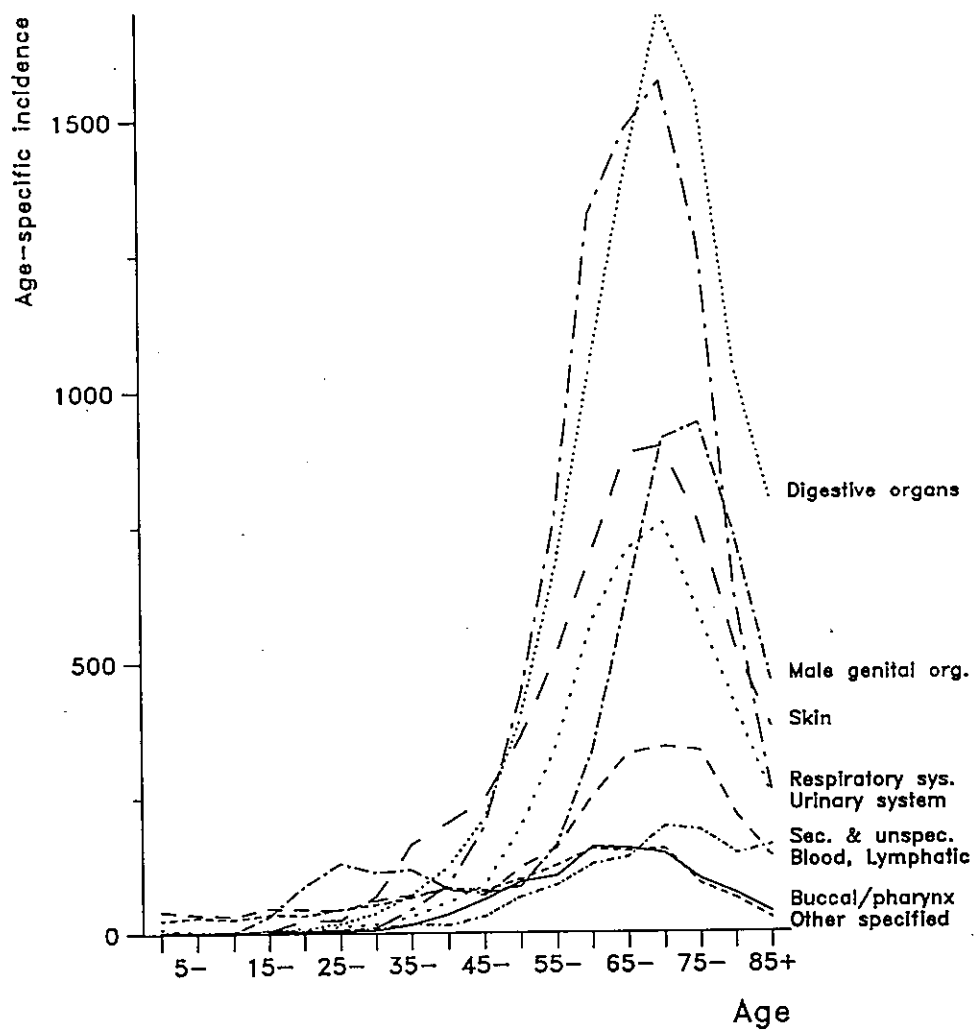


図5. 年齢階級別罹患率曲線—デンマーク、1983-85年、男、特定部位
(線グラフ)

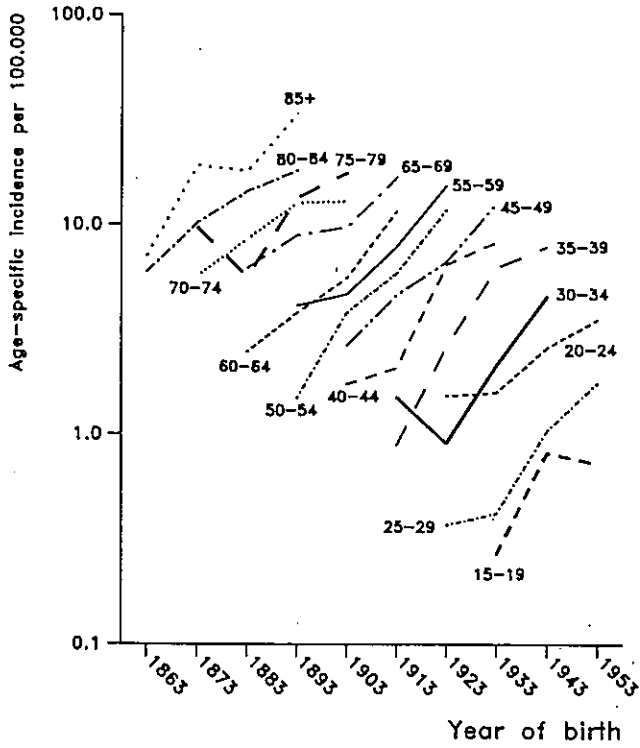


図6. 皮膚黒色腫の出生年別、年齢階級別罹患率—デンマーク、男 (線グラフ)

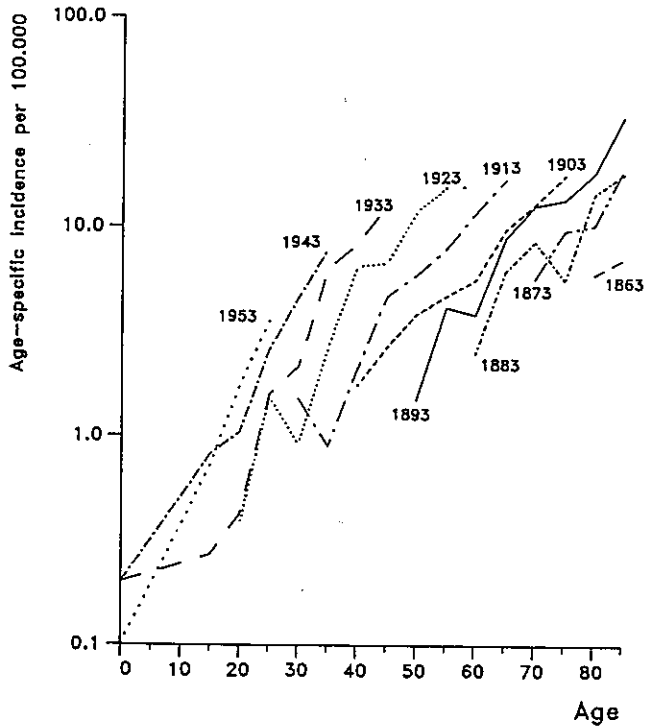


図7. 皮膚黒色腫の出生年別、年齢階級別罹患率—デンマーク、男 (線グラフ)
 訳者注：図6とは図示方法がちがっているが、同じ目的で作られている。

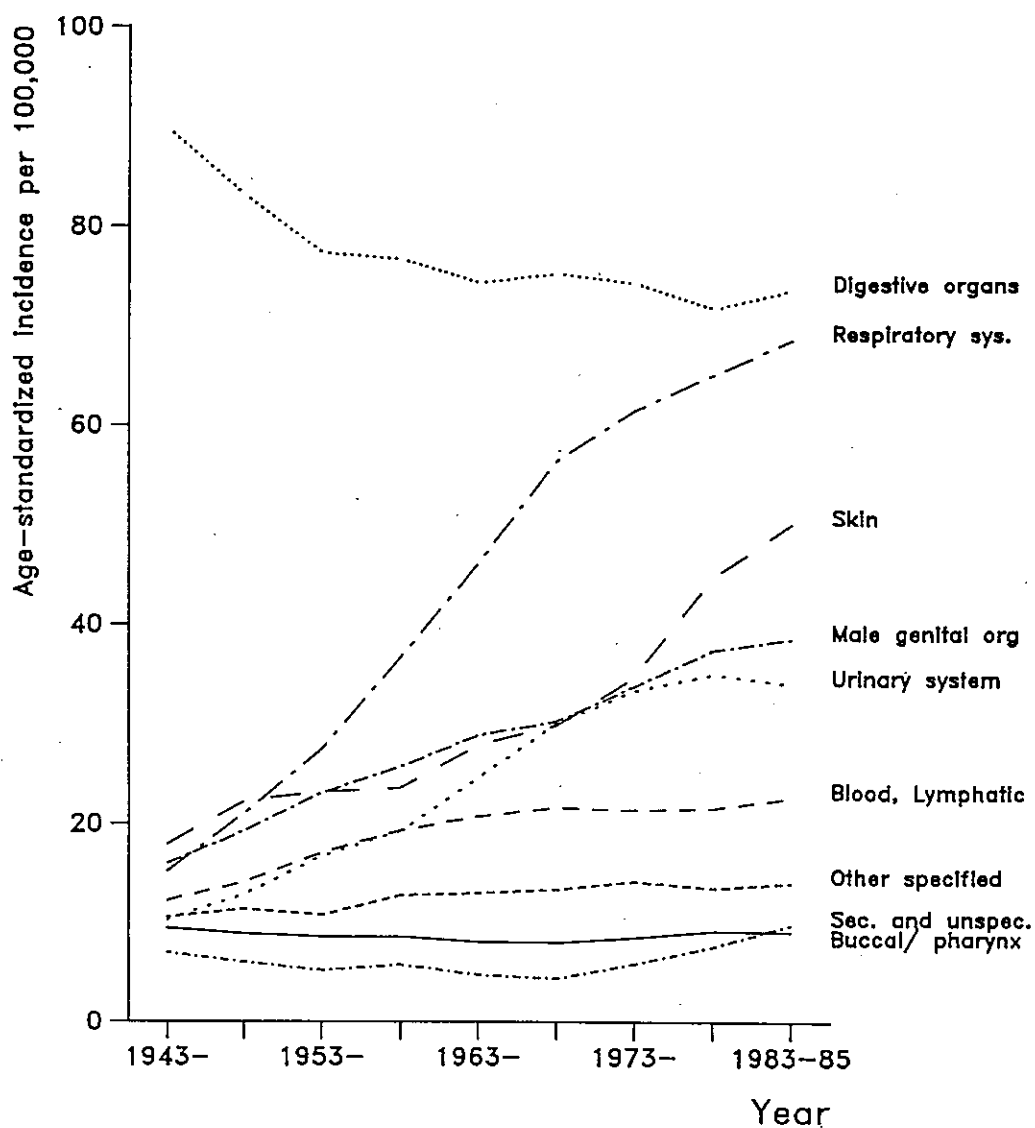


図8. 年齢調整罹患率の年次推移—デンマーク、1943-85年、男、特定部位
(線グラフ、縦軸は標準目盛)

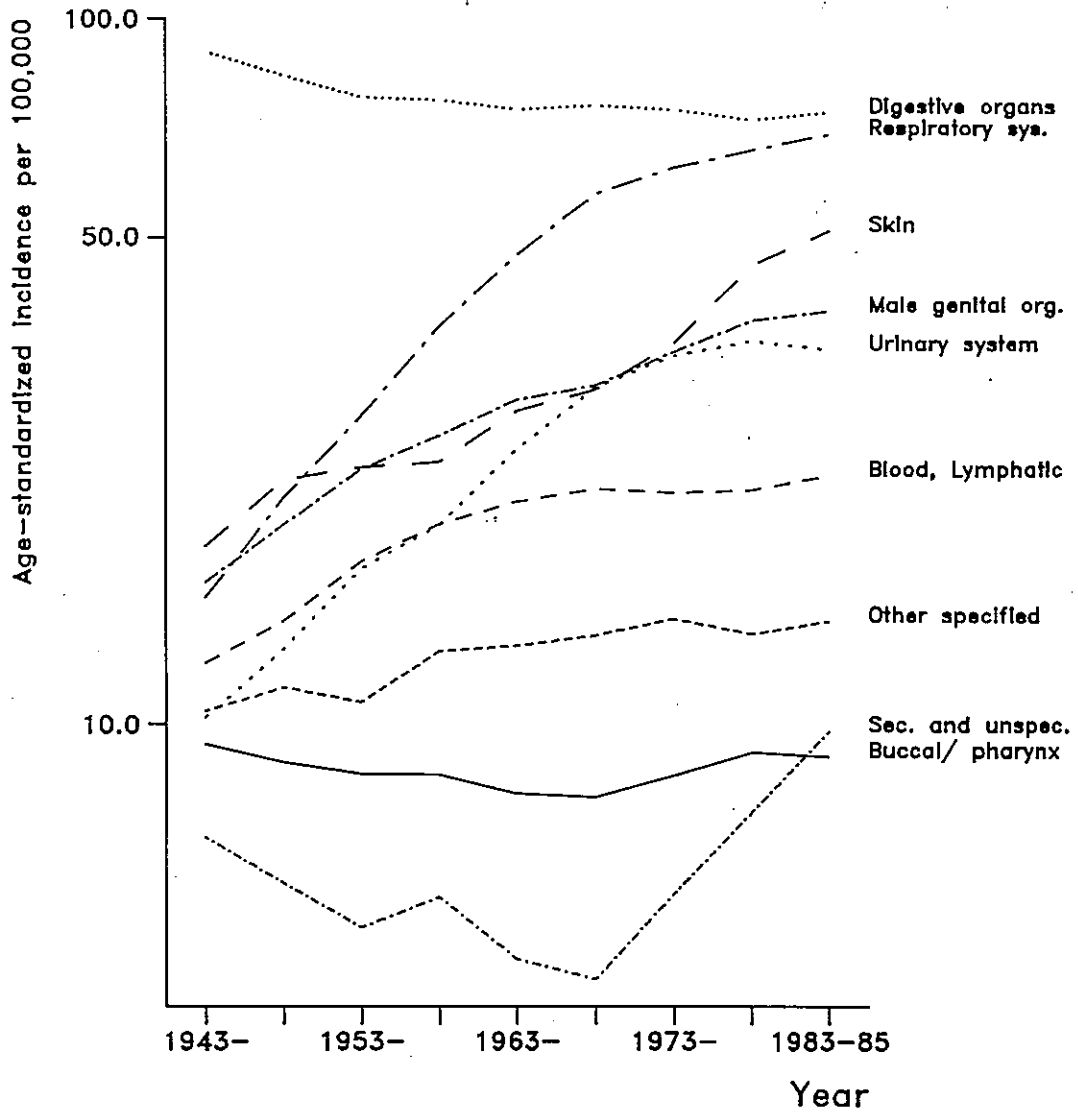


図9. 年齢調整罹患率の年次推移—デンマーク、1943-85年、男、特定部位
 (線グラフ、縦軸は対数目盛)
 (訳者注：表8の標準目盛と異なることに注意)

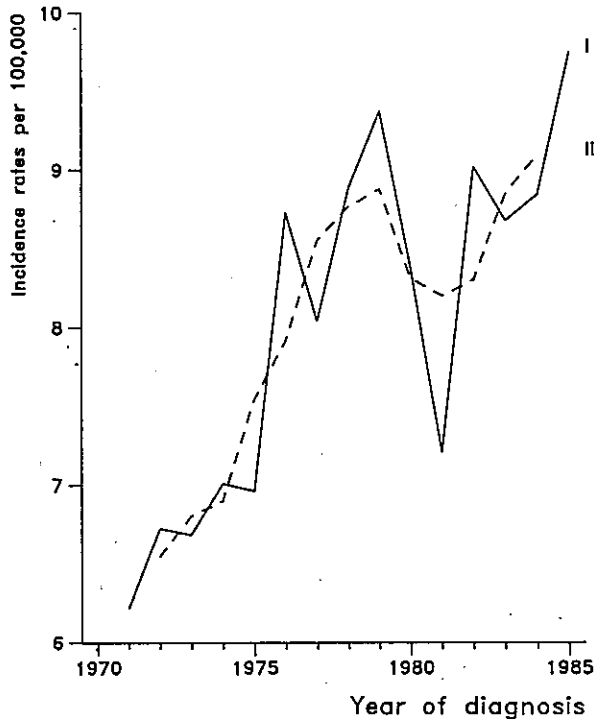


図10. 睪丸がんの年齢調整罹患率の年次推移—デンマーク、1971-85年
 I : 毎年の率 II : 3年移動平均の率

種々の要素が全体に占める寄与の程度を、円グラフによって表すことができる。円グラフは、全体に占める各々の変数の割合を示す「くさび」に分けられた円である。全円360度が100%、即ち、1%=3.6度で表される。全円は、分度器を用いて分けられる。例を図3に示す。

全体に占める構成要素の大きさを示す、もうひとつの方法は、帯グラフを用いることである。棒グラフの棒の代りに帯を使うことで、数字データや順位データの解析に利用できる。独立した群ごとの構成要素を比較する場合に特に便利である。縦でも横でも、読みやすい方向で描ける。例を図4に示す。

年齢階級別罹患率は、線グラフで表されることが最も多い。等間隔でも片対数（年齢を等分、率を対数軸にとる）でも描くことができる。対数グラフでは、率の相対的な変動は、絶対値にかかわらず、同じ傾きを持つ。年齢階級別罹患率をプロットすると、図5に示すように、部位別の年齢曲線の違いや、時期ごとの違いを簡単に表すことができる。年齢階級別罹

患率の推移を表すためにも、線グラフが最適である。図6に示すように、推移と、birth cohort（出生年別のコホート）の年齢階級別罹患率とを簡単に組み合わせて表すことができる。各々の birth cohort の年齢階級別罹患率は、図7のように示すこともできる。毎年の年齢調整罹患率は、図8のように等間隔で、また同じデータの率を対数軸でプロットすると図9のように描くことができ、増加率を部位別に比較することができる。

稀な部位では、数が少ないことにより、1年間の率の変動しやすい。3年移動平均率を計算することもできる。それにより、変動を和らげ、実際の状況を明確に提供できる。3年間の発生数を合計し、同じ3年間の人口も合計して、率を求める。順次、（1年づづらして）これを繰り返す（1年目を除き、最新年を加える）。例を図10に示す。

特別報告

がんの病因、自然史、および予後に関する多くの結果が、がん登録のデータを用いて導きだされる。さらに、がん登録室には通常、そのような解析に必要なコンピュータ設備と統計的技術とがある。従って、がん登録室が疫学的、生物統計学的研究機関としての役割を果たすのは当然である。第3章でふれたように、そのような特別な研究では、地域別や人種別の罹患率を詳細に比較し、罹患や生存の年次変動を検定することも行う。さらに、登録過程そのものやデータの信頼度を扱ったり、ある特定部位の腫瘍型の組織学的分布を詳細に検討する、などがある。

このような種類の研究は、奨励されるべきである。これらのがん登録室のモノグラムとして発行したり、あるいは科学雑誌の補足として報告することもあるが、後者では、国際的に読まれることが保証される。科学雑誌の原著として掲載されることもあり、がん登録室の仕事の質の評価を上げるのに役立つ。

謝辞

「グラフを用いた提示」の節に、図の使用を許可して戴いた米国国立がん研究所人口統計学的解析部の E. M. Shambaugh 博士に感謝する。

（翻訳：味木和喜子、花井 彩）

第11章 がん登録における統計学的手法

P. Boyle and D.M. Parkin

この章は、統計の参考書の代わりとなることを意図していない。この章の目的は、単に、がん登録の関係者が、そのデータを提示するために必要な計算について理解を援助することにある。地域がん登録では、登録データを罹患率として提示する。これらの率を比較研究に用いる場合、例えば罹患率を時期や地域によって比較する場合に必要な手法をも、同様に記載している。罹患率が計算できない場合は、がん登録室の結果は割合で提示されなければならない、そのような登録室における方法についても述べる。

第1部 罹患研究の方法

定義

罹患率

地域がん登録室の大きな関心は、がんの罹患率を計算し、各がんのリスクの登録地域間の比較研究、あるいは、登録地域内のサブグループ間での比較研究に、それを用いることである（第3章参照）。

罹患は、疾患のない特定の人口集団から新たに発生したがんの数を意味し、罹患率は一定期間に発生した数を示す。即ち：

$$\text{罹患率} = \frac{\text{新発生数}}{\text{対象集団の人口}} \quad (\text{一定期間における})$$

この指標は、病気に罹る確率、もしくはリスクを直接評価する指標であり、疫学研究でも本質的に重要である。

罹患率は、期間に関係するので、新症例の発生日を正確に定義することが必要である。がん登録室では、この発生日として罹患日（第6章の項目16）を用いる。罹患日は、必ずしもがんが実際に発生した時には対応しないが、発生日として使用しうる他の日付、例えば、症状の発現日、病院初診日、治療日等では、一致した方法で定義することが、さらに難しくなる。

観察期間

疾患の本当の瞬時リスクは、非常に短時間における罹患率、即ち瞬時率もしくは罹患力により求められる。期間が長くなるにつれて、対象人口の定義がより曖昧になり（出生、死亡、転居による）、さらに率そのものが時間により変動する可能性がある。実際は、人間のがん

は、比較的可成りな事象であり、その研究のためには大きな人口を数年にわたって観察しなければならない。罹患率は、便宜上、1年間の率（即ち、/年）で表現され、数年にわたってデータが収集された場合は、分母は観察人年に変換される。

対象人口

疫学的コホート研究では、危険因子の有無についての情報を得た比較的小さな集団をフォローする。（死亡、転居、追跡不能等により）研究群から外れる症例の存在は不可避であり、また、しばしば新しい症例が対象集団に追加される。

その結果、観察下にあり、疾病のリスクのある症例数は、期間により異なる。従って、罹患率の分母は、関係する各々の人年を総和することによって計算される。

がん登録室では通常、人口集団全体に対する罹患率を計算しているが、そのような率の分母を、各々の個人が対象人口であるか否かを知ることにより導き出すことはできない。従って、一般的に中間人口（即ち、期間の最初と最後の人口の平均）が、対象人口の近似値として用いられ、国勢調査部門から得ることができる。罹患率の推計値の分散は、率の分子となる症例数によって決定される。この理由から、数年の観察数を累積し、年間平均罹患率を計算することが通常行われている。そのような場合の分母は、人年で求められ、観察期間の各年の中間人口の総和を用いるのが理想的である。これらが利用できない場合は、人口構成に急激な変化や変動がなければ、精度が低くなるが、観察期間の1点あるいは2点で得られる人口から、人年を概算して用いることができる。種々の利用可能な人口データから人年を概算する例を表1に示す。便宜上、小さな小数を使うことを避けるために、がんの罹患率は10万人年あたりで表現される。小児がんについては、しばしば100万人あたりの率で表現される。

表1. 観察人年の計算方法（スコットランド、45-49歳、男性、1980-1984年）
利用できる種々の人口データを用いた計算方法

年	1. 各年	2. 中央年	3. 不規則な年
1980	140 800 人	—	—
1981	142 700	—	142 700 人
1982	140 600	140 600 人	—
1983	141 200	—	141 200
1984	141 500	—	—

方法1. 人年 = $140\,800 + 142\,700 + 140\,600 + 141\,200 + 141\,500 = 706\,800$

方法2. 人年 = $140\,600 \times 5 = 703\,000$

方法3. 2番目の年から4番目の年までの人口減少 = 1500; 年間の減少 = $1500/2 = 750$;
人年 = $(142\,700 + 750) + 142\,700 + (142\,700 - 750) + 141\,200 + (141\,200 - 750) = 709\,750$

推定人口から観察人年を概算した場合、率の分母にリスクのない人を少数含むことがある。幸運なことに、がんの罹患率の研究では、そのがんに罹患している生存者数は非常に少ないので、推定人口と実際の対象人口との差は非常に小さい。しかし、人口の大部分が真にリスクのない場合は、分母から除外すべきである。明確な例としては、性に特有のがんの率では、反対の性を除く、あるいは子宮がんの罹患率を計算する場合に、子宮のある女性に対してのみ計算する（中年女性の中で子宮切除術を行った人は大きな部分を占めるかもしれない）。このことは、子宮切除率の異なる期間や場所での比較を行う場合には重要となる。

率の計算

地域社会における疾患の発生を表現する多くの指標が開発されてきた。これらは Inskip らによって明解に概説されているし (Inskip *et al.*, 1983)、この問題についての優れた議論を提供する他の多くの報告がある (Armitage, 1972; Armitage & Berry, 1987; Breslow & Day, 1980, 1987; Doll & Cook, 1967; Fleiss, 1981; MacMahon & Pugh, 1970)。この章では、がん登録室で働く人達にとって一般的に最も適した方法に限り、実例を提供する。例には可能な限りスコットランドの男性の肺がん罹患データを用いた。できるだけ多くの中間結果を表示するように努めたが、全てを含めることは不可能であった。また、いくつかの中間段階を繰り返すことにより、計算と四捨五入とに用いる小数点以下の桁数により、結果が若干異なる場合がある。従って再計算を試みる読者は、最終的には同じ結果になっても、中間結果がテキストの値と異なる場合もある。

粗（全年齢）および年齢階級別率

症例数と、これに対応する母集団の観察人年が算定できる年齢階級群が A 個あると仮定する。しばしば、このグループ数は 18 ($A=18$) であり、カテゴリーとしては、0-4歳、5-9歳、10-14歳、15-19歳、..., 80-84歳、85歳以上 (85+) が用いられる。しかし、種々の分類がしばしば用いられる。例えば、小児を0歳と1-4歳とに分けるとか、分類を75歳までに短縮、即ち70-74歳と75歳以上にする場合、などがある。

r_i を i 番目の年齢階級群での症例数とする。全ての症例について年齢がわかる場合、全症例数 R は以下のように書ける。

$$R = \sum_{i=1}^A r_i = r_1 + r_2 + r_3 + \dots + r_A \quad \text{式(11.1)}$$

(以下の式番号には、式の文字を省略した)

同様に、 n_i を、症例と同じ期間の i 番目の年齢階級における観察人年とすると、全観察人年 N は以下のように書ける。

$$N = \sum_{i=1}^A n_i = n_1 + n_2 + n_3 + \cdots + n_A \quad (11.2)$$

粗罹患率、即ち全年齢の10万人あたりの率は、全症例数(R)を全観察人年(N)で割り、10万を乗じることで簡単に得ることができる。

$$\text{粗罹患率} = C = \frac{R}{N} \times 100\,000 \quad (11.3)$$

従って全例の年齢がわかっている場合

$$C = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A n_i} \times 100\,000 \quad (11.4)$$

年齢階級*i*の年齢階級別率、即ち a_i は、その年齢階級の症例数(r_i)を対応する観察人年(n_i)で割り、10万をかけることにより、人口10万に対する率として求めることができる。つまり、

$$a_i = \frac{r_i}{n_i} \times 100\,000 \quad (11.5)$$

年齢調整—総論

がんの疫学において最もよくみられる問題の一つは、異なる集団間、又は同じ集団の異なる時期でのがんの罹患率を比較することである。単なる粗罹患率を比較すると、比較する集団の人口構成の差によって、間違った結論が導きだされることが、しばしばある。もし一方の集団が他方より平均的に若ければ、たとえ年齢階級別率が両者で同じであっても、人口の若い方より年寄りの方で多くの症例が発生するであろう。表2をみると、年齢の上昇に伴い、年齢階級別率が急激に増加することがわかる。

従って、2つの地域でがんのレベルを比較する場合や、同一地域でのがんの時間的変動を調べる場合には、年齢構成の変化や違いを考慮することが重要である。これは、年齢調整を行うことにより達成できる。しかし、異なった年齢構成を持つ集団での率を比較する場合の困難は、比較が各々の年齢階級別率に限られる場合に限り、完全に乗り越えることができるということを、強調しなければならない。この点は、いくら強調してもしたりない。年齢調整のような技術で計算された要約指標は、年齢階級別率の吟味にとってかわることはできない。しかし、年齢調整率のような要約指標が利用できることは、特に多くの罹患率を比較する場合に便利である。

例1. 粗率および年齢階級別率の計算

表2は、スコットランド男性の気管、気管支、肺のがん（国際疾病分類（ICD-9）162）の罹患に関するデータを示している。症例数と人口は、1980年から1984年までの総和である。

表2. 肺がん罹患数および率—スコットランド、男、1980-84年

年齢階級 番号 i	年齢階級	症例数 r_i	観察人年 n_i	年齢階級別率 対10万人 r_i/n_i
1	0-4 歳	0 人	827 400	0.00
2	5-9	0	856 500	0.00
3	10-14	0	1 061 500	0.00
4	15-19	0	1 157 400	0.00
5	20-24	4	1 074 900	0.37
6	25-29	3	917 700	0.33
7	30-34	29	890 300	3.26
8	35-39	61	816 000	7.48
9	40-44	153	724 400	21.12
10	45-49	376	706 800	53.20
11	50-54	902	703 800	128.16
12	55-59	1 819	691 200	263.17
13	60-64	2 581	610 900	422.49
14	65-69	3 071	511 800	600.04
15	70-74	3 322	425 600	780.55
16	75-79	2 452	266 800	919.04
17	80-84	1 202	122 500	981.22
18	85+	429	54 700	784.28
	全年齢	16 404	12 420 200	

年齢階級別率は、式 (11.5) を用いて計算できる。例えば、40-44歳 ($i=9$) では、

$$a_9 = \frac{r_9}{n_9} \times 100\,000 = \frac{153}{724\,400} \times 100\,000 = 21.1$$

従って、40-44歳では、平均の年間年齢階級別罹患率は、人口10万人あたり21.1となる。同様に計算した他の年齢階級別率を表2に示す。

粗率 C は、式 (11.3) を用いて、計算できる。全症例数 R は16 404、全観察人年 N は12 420 200であるので、

$$C = \frac{16\,404}{12\,420\,200} \times 100\,000 = 132.1$$

従って、スコットランド男性の1980-84年の平均の年間肺がん罹患率は、人口10万人あたり132.1である。

一般的に用いられている年齢調整方法には2種、即ち、直接法と間接法とがある。まず直接法について述べる。直接法は、間接法に比べて解釈上の利点が相当あり（詳細については例えば Rothman, 1986を参照）、一般的にいうと、可能な限り直接法を用いるべきである。〔詳細は、Breslow and Day (1987) の報告（72-75頁）を参照のこと。〕

年齢調整—直接法 (Age-standardization — direct method)

年齢調整率は、観察集団の年齢階級別率が、ある参照人口に適応された場合に発生する理論上の率である。この参照人口は、通常、標準人口と言われている。

標準人口の各々の年齢階級の人口は、標準化で用いられる重みとして知られている。この重み w_i として、種々のセットが使用可能である。異なる重みを用いると（即ち、異なる標準人口を用いると）、異なる年齢調整率が得られる。最も頻繁に用いられているのは世界人口（表3）であり、Segi (1960)によって提案され、Dollら(1966)が改訂した人口である。この世界人口は「5大陸のがん罹患」シリーズでも用いられている。この世界人口が広く用いられていることが、地域間のがんのレベルの比較に役立っている。

標準人口の*i*番目の年齢階級の人口を w_i とし、前述と同様に、 $i = 1, 2, \dots, A$ で、 a_i を*i*番目の年齢階級での年齢階級別率とすると、年齢調整率 (Age-standardized rate; ASR) は以下のように計算できる。

表3. 世界標準人口 (Dollら、1966)

年齢階級番号 (<i>i</i>)	年齢階級	人口 (w_i)
1	0-4 歳	12 000 人
2	5-9	10 000
3	10-14	9 000
4	15-19	9 000
5	20-24	8 000
6	25-29	8 000
7	30-34	6 000
8	35-39	6 000
9	40-44	6 000
10	45-49	6 000
11	50-54	5 000
12	55-59	4 000
13	60-64	4 000
14	65-69	3 000
15	70-74	2 000
16	75-79	1 000
17	80-84	500
18	85+	500
全年齢		100 000

例2. 直接法による年齢調整率の計算

表4に、この章で記述したスコットランド男性の肺がんの年齢階級別率を再掲する。この年齢階級別率 (a_i) に、標準人口 (w_i) を乗じると、 $a_i w_i$ が得られ、これを合計すると9 062 410となる。即ち、

$$\sum_{i=1}^A a_i w_i = 9\,062\,410$$

表4. 年齢調整肺がん罹患率—直接法—スコットランド、男性、1980-84年

年齢階級 番号 i	年齢階級	年齢階級別率 対10万人 a_i	世界標準 人口 w_i	$a_i \times w_i$
1	0-4	0.00	12 000	0
2	5-9	0.00	10 000	0
3	10-14	0.00	9 000	0
4	15-19	0.00	9 000	0
5	20-24	0.37	8 000	2 960
6	25-29	0.33	8 000	2 640
7	30-34	3.26	6 000	19 560
8	35-39	7.48	6000	44 880
9	40-44	21.12	6000	126 720
10	45-49	53.20	6000	319 200
11	50-54	128.16	5000	640 800
12	55-59	263.17	4000	1 052 680
13	60-64	422.49	4000	1 689 960
14	65-69	600.04	3 000	1 800 120
15	70-74	780.55	2 000	1 561 100
16	75-79	919.04	1 000	919 040
17	80-84	981.22	500	490 610
18	85+	784.28	500	392 140
計			100000	9 062 410

利用した標準人口は、世界人口であり、それらを合計すると便宜上、100 000になる。即ち、

$$\sum w_i = 100\,000$$

スコットランド男性の1980-84年の人口10万人あたりの肺がん平均年間年齢調整罹患率 (ASR) は、以下のように、計算する。

$$ASR = \frac{\sum_{i=1}^A a_i w_i}{\sum_{i=1}^A w_i} = \frac{9\,062\,410}{100\,000} = 90.62410$$

即ち、人口10万人年あたり90.6である。(ASRの単位は10万人年あたりであり、計算に用いた年齢階級別率 a_i の単位でもある。)

$$ASR = \frac{\sum_{i=1}^A a_i w_i}{\sum_{i=1}^A w_i} \quad (11.6)$$

年齢不詳のがん症例が含まれている場合がある。このことは、式(11.1)がもはや不当であることを意味する。なぜなら、全症例数(R)は、各年齢階級別の症例数の総和($\sum r_i$)よりも大きくなるからである。従って、ASRも、年齢階級別率(式11.5)に由来するので、真実値よりも低めに表現することになる。

Doll & Smith (1982) は、ASRに以下を乗じることによって、補正することを提言している。

$$\frac{R}{\sum_{i=1}^A r_i}$$

この補正を用いるための前提条件は、年齢不詳の症例の年齢分布が、年齢のわかっている症例の年齢分布に等しいことである。年齢が記録されていない症例は年長者に多いので、この仮定は不当であることが多いが、年齢不詳の割合が小さい限り(5%未満)、この影響は通常大きくない。

区間率 (Truncated rate)

Doll & Cook (1967) は、年齢階級を35-64歳に限った率の計算を提案している。この理由は、主として、老人ではがんの診断と登録が不確かであることが多く、老人での年齢階級別率の正確性に疑いがあるからである。この方法で正確さが多少増加するとはいえ、計算と解釈とが幾分難しくなることや、収集したデータの多くを無駄にすることに見合うだけの利点があるかどうかは、議論の余地のあることであるが、この区間率を用いてデータを提示し続けている研究者もいる。実際、計算方法は、表4のデータの一部を用いるだけである。

区間年齢調整率 (Truncated age-standardized rate; TASR) は、以下のように書ける。

$$TASR = \frac{\sum_{i=8}^{13} a_i w_i}{\sum_{i=8}^{13} w_i} \quad (11.7)$$

明らかに、式(11.7)は、式(11.6)の特別な例、即ち年齢階級を8番目(35-39歳)から13番目(60-64)に限った値である。同様に、小児がんの罹患率を比較する場合には、0-14歳の区間年齢が、標準人口の対応する部分とともに、用いられている (Parkin *et al.*, 1988)。

例3. 直接法による区間年齢調整率の計算

表5に、表4のうち区間年齢調整率の計算に関係する部分を示した。年齢の区間は35歳から64歳である。

表5. 区間(35-64)年齢調整肺癌罹患率—直接法—スコットランド、男性、1980-84年

年齢階級 番号 i	年齢階級	年齢階級別率 対10万人 a _i	世界標準 人口 w _i	a _i × w _i
1	0-4			
2	5-9			
3	10-14			
4	15-19			
5	20-24			
6	25-29			
7	30-34			
8	35-39	7.48	6000	44 880
9	40-44	21.12	6000	126 720
10	45-49	53.20	6000	319 200
11	50-54	128.16	5000	640 800
12	55-59	263.17	4000	1 052 680
13	60-64	422.49	4000	1 689 960
14	65-69			
15	70-74			
16	75-79			
17	80-84			
18	85+			
計			31000	3 874 240

この例での注意点は

$$\sum_{i=8}^{13} w_i = 31\,000$$

さらに、

$$\sum_{i=8}^{13} a_i w_i = 3\,874\,240$$

重み (w_i) は、a_iw_iと同じ区間で合計することが必須である。従って、

$$\text{TASR} = \frac{\sum_{i=8}^{13} a_i w_i}{\sum_{i=8}^{13} w_i} = \frac{3\,874\,240}{31\,000} = 124.97548$$

即ち、スコットランド男性の1980-84年の肺癌の平均の年間区間(35-64)年齢調整罹患率は、人口10万人あたり125.0である。

年齢調整率の標準誤差—直接法

実際の値から計算された年齢調整罹患率は、統計学上は、ある種の真のパラメーター値の概算値（真の値は、観察群が十分に多い時に、始めて知りうる）として考えるべきである。従って、概算率の精度を示す指標、例えば率の標準誤差を提示することが普通である。

標準誤差は、率の信頼区間を求めるためにも用いられる。信頼区間は、直観的で解釈しやすい。95%信頼区間は、真の罹患率が95%の確率で、その範囲内に存在する（即ち、概算値100個のうち5個だけが真の値を含まない信頼区間をもつ）ということである。代わりに99%信頼区間を用いることもあるが、信頼度を高めるために、95%信頼区間よりも広い区間となる。

一般的に、年齢調整率 (ASR) の $(100-(1-\alpha))\%$ 信頼区間は、標準誤差 $s.e.(ASR)$ を用いると、以下のように表される。

$$ASR \pm Z_{\alpha/2} \times (s.e.(ASR)) \quad (11.8)$$

$Z_{\alpha/2}$ は標準正規偏差を意味する（基本概念については Armitage and Berry(1987) を参照）。例えば、95%信頼区間は、正規分布の97.5パーセントイル値1.96を $Z_{\alpha/2}$ として選択することにより計算できる。99%信頼区間では、 $Z_{\alpha/2}$ は2.58となる。

直接法による年齢調整率の標準誤差を求める方法として、2つの計算方法がある。2項分布とポアソン近似である。例を以下に示す。両者ともに同じような結果となるので、どちらを用いてもよい。

年齢調整罹患率 (ASR) は式 (11.6) から計算できる。ASRの分散は、以下のように表される。

$$\text{Var}(ASR) = \frac{\sum_{i=1}^A [a_i w_i^2 (100\,000 - a_i) / n_i]}{\left(\sum_{i=1}^A w_i \right)^2} \quad (11.9)$$

ASRの標準誤差($s.e.(ASR)$)は、以下の様に簡単に計算できる。

$$s.e.(ASR) = \sqrt{\text{Var}(ASR)} \quad (11.10)$$

例2で示したASRの95%信頼区間は、式(11.8)から求められる。

$$\begin{aligned} ASR \pm Z_{\alpha/2} \times (s.e.(ASR)) &= 90.62 \pm (1.96 \times 0.73) \\ &= 89.12 \sim 92.05 \end{aligned}$$

Armitage and Berry (1987)が示すように、 a_i が小さい場合（多くの場合）、 a_i の2項分布をポアソン近似にかえて、求めることができる。この結果、ASRの分散($\text{Var}(ASR)$)は、以下の式で表される。

$$\text{Var}(ASR) = \frac{\sum_{i=1}^A [a_i w_i^2 \times 100\,000 / n_i]}{\left(\sum_{i=1}^A w_i \right)^2} \quad (11.11)$$

ASRの標準誤差は、前述（式11.10）のように、分散の平方根である。

例4. 年齢調整率の標準誤差の計算（2項近似）

表6に、スコットランド男性の1980-84年の肺がんの年齢調整罹患率の標準誤差を計算するためのデータを示す。

表6. 肺がんの平均年間年齢調整罹患率の標準誤差の計算（2項分布法による）
—スコットランド、男性、1980-84年

年齢階級	年齢階級別率 対10万人 a_i	世界標準 人口 w_i	観察人年 n_i	$a_i w_i^2 (100\,000 - a_i) / n_i$
0-4	0.00	12000	827 400	0
5-9	0.00	10000	856 500	0
10-14	0.00	9000	1 061 500	0
15-19	0.00	9000	1 157 400	0
20-24	0.37	8000	1 074 900	2 202 988
25-29	0.33	8000	917 700	2 301 398
30-34	3.26	6000	890 300	13 181 644
35-39	7.48	6000	816 000	32 997 532
40-44	21.12	6000	724 400	104 936 424
45-49	53.20	6000	706 800	270 823 584
50-54	128.16	5000	703 800	454 659 552
55-59	263.17	4000	691 200	607 586 624
60-64	422.49	4000	610 900	1 101 862 784
65-69	600.04	3000	511 800	1 048 838 592
70-74	780.55	2000	425 600	727 873 536
75-79	919.04	1000	266 800	341 301 952
80-84	981.22	500	122 500	198 284 096
85+	784.28	500	54 700	355 634 848

$$\sum [a_i w_i^2 (100\,000 - a_i) / n_i] = 5\,262\,486\,016$$

さらに、 $(\sum w_i)^2 = 10\,000\,000\,000$

従って、式(11.9)から

$$\begin{aligned} \text{Var (ASR)} &= 5\,262\,486\,016 / 10\,000\,000\,000 \\ &= 0.526249 \end{aligned}$$

さらに、式(11.10)から、 $s.e. (ASR) = 0.73$

従って、スコットランド男性の1980-84年の肺がんの平均年間年齢調整罹患率の標準誤差は0.73である。

例 5. 年齢調整率の標準誤差の計算 (ポアソン近似)

表 7 に、この 2 つめの方法による年齢調整率の標準誤差を計算するためのデータを示す。

表 7. 年齢調整率の標準誤差 (ポアソン近似)

年齢階級	年齢階級別率 対10万人	世界標準 人口	観察人年	$\frac{a_i w_i^2 \times 100\,000}{n_i}$
	a_i	w_i	n_i	
0-4	0.00	12000	827 400	0
5-9	0.00	10000	856 500	0
10-14	0.00	9000	1 061 500	0
15-19	0.00	9000	1 157 400	0
20-24	0.37	8000	1 074 900	2 202 996
25-29	0.33	8000	917 700	2 301 406
30-34	3.26	6000	890 300	13 182 074
35-39	7.48	6000	816 000	33 000 000
40-44	21.12	6000	724 400	104 958 592
45-49	53.20	6000	706 800	270 967 744
50-54	128.16	5000	703 800	455 242 967
55-59	263.17	4000	691 200	609 189 824
60-64	422.49	4000	610 900	1 106 537 856
65-69	600.04	3000	511 800	1 055 169 984
70-74	780.55	2000	425 600	733 599 616
75-79	919.04	1000	266 800	344 467 776
80-84	981.22	500	122 500	200 248 976
85+	784.28	500	54 700	358 446 048

$$\sum (a_i w_i^2 \times 100\,000 / n_i) = 5\,289\,515\,520$$

さらに、 $(\sum w_i)^2 = 10\,000\,000\,000$

従って、式 (11.11) から

$$\begin{aligned} \text{Var (ASR)} &= 5\,289\,515\,520 / 10\,000\,000\,000 \\ &= 0.52895 \end{aligned}$$

さらに、式 (11.10) から、 $s.e. (\text{ASR}) = 0.73$

この例では、前述の方法によるものと、小数点以下 2 桁では結果は同じであり、2 つの方法が似ていることを示している。

直接法で計算した2つの年齢調整率の比較

例えば、異なる地域、人種、期間のような、異なる集団間の直接法による年齢調整率の比を調べることは、しばしば興味の対象となる。直接法による年齢調整率の比、 ASR_1/ASR_2 は、標準化率比(standardized rate ratio; SRR)と呼ばれ、集団2に対する集団1のその疾患の相対危険度(relative risk)を意味する。標準化率比の統計的有意性の検定(観察された比が有意に異なっているか否かの指標)も通常行われている。標準化率比の正確な信頼区間を求めるためには、種々の計算方法がある(Breslow & Day, 1987(p.64); Rothman, 1986; Checkoway *et al.*, 1989)。近似値は以下の式で求められる(Smith, 1987)。

$$= (ASR_1/ASR_2)^{1 \pm (Z_{\alpha/2}/X)} \quad (11.12)$$

但し

$$X = \frac{(ASR_1 - ASR_2)}{\sqrt{(s.e.(ASR_1))^2 + s.e.(ASR_2)^2}}$$

$Z_{\alpha/2} = 1.96$ (95%レベル)、もしくは $Z_{\alpha/2} = 2.58$ (99%レベル)

もし、この信頼区間が 1.0 を含むならば、 ASR_1 と ASR_2 との間に統計的な差はない ($Z_{\alpha/2} = 1.96$ ならば 5% レベル、 $Z_{\alpha/2} = 2.58$ ならば 1% レベル)。

例6. 標準化率比の信頼区間の計算

スコットランド男性の1980-84年の肺がんの年齢調整率は90.6、で標準誤差は0.73であった。1960-64年では、それに該当する率は68.3で、標準誤差は0.67であった。

標準化率比 $ASR_1/ASR_2 = 90.6/68.3 = 1.3265$

標準化率比の信頼区間を求めるために、式(11.12)を用いると

$$X = \frac{(ASR_1 - ASR_2)}{\sqrt{(s.e.(ASR_1))^2 + s.e.(ASR_2)^2}} = \frac{90.6 - 68.3}{\sqrt{(0.73)^2 + (0.67)^2}} = \frac{22.3}{\sqrt{0.9818}} = 22.51$$

$Z_{\alpha/2} = 1.96$ とすると、95%信頼区間を求められる。従って、

$$\text{下限値} = (ASR_1/ASR_2)^{1 - (Z_{\alpha/2}/X)} = (1.3265)^{1 - (1.96/22.51)} = 1.29$$

$$\text{上限値} = (ASR_1/ASR_2)^{1 + (Z_{\alpha/2}/X)} = (1.3265)^{1 + (1.96/22.51)} = 1.36$$

もし、2つの期間の率が同じであれば、比(ASR_1/ASR_2)は1.0となる。しかし、この比の95%信頼区間(1.29, 1.36)は、この値を含まないので、比は5%のレベルで有意差があると結論できる。

集団のサブグループ間の調整率を比較する場合は、考えうる全ての組み合わせで比較するよりも、各々の率を集団全体の率と比較する方が論理的である。例えば、スコットランドのがん罹患アトラスの作成にあたりKenpら(1985)は、1975-80年の6年間について、56地区の分子と分母の情報を入手した。部位別、性別に10万人年あたりの平均罹患率を、世界人口(前述)を標準として直接法で計算した。同様に、各々の地区とスコットランド全体での標準誤差も計算した。人口の多い地域(例えばGlasgowはスコットランドの全人口の17%を占める)を比較する場合の影響を除くために、即ち、スコットランド全体はその地区の影響を受けているので、各地区の率と残りのスコットランドの率とを比較した(例えば、Glasgow対Glasgow以外のスコットランド)。比較の方法は、前述の直接法による年齢調整率の比較方法を用い、比を以下のように示した。即ち、1%危険率で高い(++)、5%危険率で高い(+)、有意差なし、5%危険率で低い(-)、1%危険率で低い(--)。

表8に、スコットランドがん地図(Kempら, 1985)から抜粋した肺がんの罹患率を示す。男性では、地区33、Glasgow市の率が最も高く(10万人あたり130.6、標準誤差2.01)、1%の危険率で、残りのスコットランドより高かった。隣接するInverclyde(109.9, 5.35)も、1%の危険率で有意に高く、Edinburgh市(103.2, 2.32)も同様であった。人口の大きさが統計的有意レベルに与える影響は、留意すべきである。Edinburgh市は、男性の肺がん罹患率では7番目であるが、人口が多いので、統計的有意性の高い3地区のひとつになっている。

同様の傾向が女性にも認められ、Glasgow市(33.3, 0.90)が最も高い。しかし、2番目に高いBadenoch(31.8, 9.36)は、人口が少ないために、残りのスコットランドとの差がみられなかった。

表8. スコットランドの幾つかの地区における肺がん罹患率, 1975-80年 (Kempら, 1985年)

地区		男				女			
番号	地域名	症例数	ASR	SE	順位	症例数	ASR	SE	順位
7	Badenoch	21	55.0 --	12.58	47	14	31.8	9.36	2
21	Edinburgh	2087	103.2 ++	2.32	7	734	25.9 ++	1.05	8
24	Tweeddale	49	73.6	11.03	28	28	29.1	6.24	3
33	Glasgow	4579	130.6 +++	2.01	1	1802	33.3 ++	0.90	1
37	Cumbernauld	145	109.1	9.25	3	36	21.8	3.58	18
45	Inverclyde	438	109.9 ++	5.35	2	137	27.5	2.46	5
54	Orkney	36	40.2 --	6.97	56	12	13.6 -	4.07	48
55	Shetland	39	46.1	7.70	53	7	5.8 --	2.33	56
全スコットランド		19 239	91.4	0.67		6136	23.1	0.31	

ASR : 人口10万人あたりの年齢調整率 (直接法、世界標準人口)

SE : 標準誤差

++ : 残りのスコットランドより有意に高い (p<0.01)

-- : 残りのスコットランドより有意に低い (p<0.01)

- : 残りのスコットランドより有意に低い (p<0.05)

年齢調整率の傾向の検定

前述した2つの年齢調整率の比較の検定の応用として、何らかの尺度の順番に並んだ集団の率が、利用できる場合がある。この尺度のカテゴリーは、曝露の程度であったり、病因であったり、単なる時間であったりする。簡単な例としては、期間別や社会経済階層別の年齢調整率がある。地区別（例えば県別）の率を、例えば平均降雨量や、標高、大気汚染度、等の順番に並べかえることもできる。

このような状況では、研究者は2つの年齢調整率を比較するだけでなく、曝露の程度に応じて、罹患率にも何らかの傾向があるかどうかを調べることは、興味のあることであろう。回帰直線を当てはめるのが、直線傾向を表すもっとも簡単な方法である。

例として、1960-70年のスコットランド男性の年齢調整肺がん年間罹患率を用いる。時間的な傾向を評価するために、率を並べるのに実際の年を用いる。しかし、計算を簡単にするために、各年から1959を引いて、1960年を1、1961年を2、1970年を11としてもよい。どちらの変数を用いても、傾向に関して同じ結果が得られる。

単回帰では(注)、2種類の変数がある。説明変数（この場合は年、 x で示す）と従属変数（この場合は年齢調整率、 y で示す）である。直線回帰式は以下のように書ける。

$$y = a + bx \quad (11.13)$$

但し

y = 年齢調整肺がん罹患率

x = 年番号（西暦から1959を引く）

a = 切片

b = 傾き

a, b の計算式と各々の標準誤差に関しては、Bland (1987) 参照のこと。例えば、

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

この式は以下のようにも書ける。

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad (11.14)$$

注：重み付きの回帰式の方が、より適切な場合もある。その場合、回帰直線を得るために各点が寄与する情報量は、均等でない。通常、 $w_i = 1 / \text{Var}(y_i)$ という重みを用いる。Armitage and Berry (1987) を参照すること。

但し

$$n = \text{観察ペア数}, y = \sum y / n, x = \sum x / n$$

傾き b の標準誤差は

$$\text{s.e.}(b) = \sqrt{\frac{1}{n-2} \left\{ \sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2 \right\}} \quad (11.15)$$

切片 a は以下の式で計算できる。

$$a = \bar{y} - b\bar{x} \quad (11.16)$$

求められた傾き (b) は、説明変数が 1 あがるごとの年齢調整罹患率の増加の平均、即ちこの例では、一年ごとの平均増加を示す。傾きの標準誤差 ($\text{s.e.}(b)$) を用いて、式 (11.8) と同じ式から、傾きの信頼区間を求めることができる。

傾きが 0 から有意に異なっているか否かの公式な検定は、傾きとその標準誤差との比 ($b / \text{s.e.}(b)$) を求め、それが自由度 $n-2$ の t 分布に従うことを用いて行うことができる (詳細は Armitage and Berry (1987) を参照)。

年齢調整—間接法

年齢調整によく用いられる別の方法として、間接法による年齢調整がある。この方法は、観察値と期待値との比較と考えるとよい。期待値は、標準となる年齢階級別率を対象集団の人口に掛け合わせることで求められる。即ち、

$$\sum_{i=1}^A e_i = \sum_{i=1}^A a_i n_i / 100\,000 \quad (11.17)$$

e_i は、年齢階級 i で期待される症例数であり、その年齢階級 i の標準率と対象集団の人口とを掛けたものである。

標準化比 (M) は、観測値 ($\sum r_i$) を期待値 ($\sum e_i$) と比較することで計算できる。

$$M = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A e_i} = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A a_i n_i / 100\,000} \quad (11.18)$$

この値は通常 100 を掛けてパーセントで表される。罹患データに適應される場合は、標準化罹患比 (Standardized incidence ratio; SIR)、死亡データに適應される場合は標準化死亡比 (Standardized mortality ratio; SMR) という。

例7. スコットランド男性の肺がんの年齢調整罹患率における平均の年間の変化の計算および傾向の有意性の検定

1960年から1970年の間では、スコットランド男性の肺がんの年間の全年齢の罹患率は、10万人あたり、それぞれ77.05、81.78、89.05、85.68、87.04、89.97、100.50、104.85、104.77、および107.57であった。

この例では、説明変数 (x) は、(年-1959) である。換言すると、1960年は1、1961年は2、同様に1970年は11となる。

$$n = 11$$

$$\sum x_i = 66$$

$$\sum x_i^2 = 506$$

$$\sum y_i = 1016.0$$

$$\sum y_i^2 = 94913.0$$

$$\sum x_i y_i = 6419.2$$

従って、式 (11.14) から

$$b = \frac{6419.2 - ((66 \times 1016) / 11)}{506 - ((66 \times 66) / 11)} = \frac{323.2}{100} = 2.938$$

さらに、

$$\bar{x} = \sum x_i / n = 6$$

$$\bar{y} = \sum y_i / n = 92.36$$

従って、式 (11.15) から $s.e. (b) = 0.351$

傾き (b) の95%信頼区間は、下記のように計算できる。

$$\text{下限} = 2.938 - (1.96 \times 0.351) = (2.938 - 0.688) = 2.250$$

$$\text{上限} = 2.938 + (1.96 \times 0.351) = 3.626$$

従って、b の95%信頼区間は (2.25, 3.63) である。傾きb が、0と有意に異なるか否かの検定をするためには、次の値を計算する

$$b / s.e. (b) = 2.938 / 0.351 = 8.375$$

これは、自由度 (n-2) の t 分布の臨界値である。ここでは、切片は1%のレベル ($p < 0.01$) で有意であり、非常に意味がある。

従って、スコットランドの男性の肺がんの罹患は、1960-1970年に有意に増加し、年齢調整率は年間平均10万人あたり約2.9人増加している。

例 8. 間接法による標準化罹患比の計算

表 9 に、1960-64年を標準とした時のスコットランドの1980-84年の男の肺がんの SIR を計算するデータを示す。

表 9. 1960-64年を標準とした時の1980-84年の肺がんの年齢調整罹患比の計算
—スコットランド、男

年齢階級	1960-64年の率 対10万人 a_i	観察人年 n_i	罹患の期待数 $e_i = a_i n_i / 100\,000$	症例数 r_i
0-4	0.00	827 400	0.00	0
5-9	0.00	856 500	0.00	0
10-14	0.00	1 061 500	0.00	0
15-19	0.09	1 157 400	1.04	0
20-24	1.12	1 074 900	12.04	4
25-29	1.70	917 700	15.60	3
30-34	4.91	890 300	43.71	29
35-39	16.25	816 000	132.60	61
40-44	29.38	724 400	212.83	153
45-49	79.92	706 800	564.87	376
50-54	151.07	703 800	1063.23	902
55-59	269.58	691 200	1863.34	1 819
60-64	391.41	610 900	2391.12	2 581
65-69	459.74	511 800	2352.95	3 071
70-74	400.46	425 600	1704.36	3 322
75-79	285.21	266 800	760.94	2 452
80-84	207.49	122 500	254.18	1 202
85+	100.84	54 700	55.16	429
		合計	11427.97	16 404

$$\sum e_i = \sum a_i n_i / 100\,000 = 11\,427.97$$

さらに、 $\sum r_i = 16\,404$

従って、1980-84年の標準化罹患比 $M \times 100$ は、式 (11.18) から

$$\frac{16\,404}{11\,427.97} \times 100 = 144$$

換言すると、男性の肺がんは、年齢構成の違いを考慮すると、1980-84年では、1960-64年より44%高い。

標準化比の標準誤差

標準化比 (M) は式 (11.18) から得ることができる。その分散 (Var(M)) は以下のとおりである。

$$\text{Var}(M) = \frac{\sum_{i=1}^A r_i}{\left(\sum_{i=1}^A a_i n_i / 100\,000\right)^2} \quad (11.19)$$

間接比の標準誤差、s.e.(M) は、前述 (式11.10) のように分散の平方根である。

$$\text{s.e.}(M) = \frac{\sqrt{\sum_{i=1}^A r_i}}{\sum_{i=1}^A a_i n_i / 100\,000} \quad (11.20)$$

Vandenbroucke (1982) は、標準化比の (100(1-α)) % 信頼区間の簡易計算法を提案した。この方法は2段階からなる。まず、事象の観察数の値を計算する。

$$\begin{aligned} \text{下限} &= [\sqrt{\text{観察数}} - (Z_{\alpha/2} \times 0.5)]^2 \\ \text{上限} &= [\sqrt{\text{観察数}} + (Z_{\alpha/2} \times 0.5)]^2 \end{aligned}$$

これらの限界値を期待値で割った値が、SIRの95% (99%) 信頼区間の近似値となる。

$$\text{SIRの下限} = \frac{[\sqrt{\text{観察数}} - (Z_{\alpha/2} \times 0.5)]^2}{\text{期待数}} \quad (11.21)$$

$$= \frac{\left\{ \sqrt{\sum_{i=1}^A r_i} - (Z_{\alpha/2} \times 0.5) \right\}^2}{\sum_{i=1}^A a_i n_i / 100\,000}$$

$$\text{SIRの上限} = \frac{\left\{ \sqrt{\sum_{i=1}^A r_i} + (Z_{\alpha/2} \times 0.5) \right\}^2}{\sum_{i=1}^A a_i n_i / 100\,000} \quad (11.22)$$

例9. 間接的に標準化した罹患比の標準誤差の計算

表9に、1960-64年に対する1980-84年のスコットランド男性の標準化罹患比(SIR)の標準誤差の計算のためのデータを示した。

$SIR = M \times 100$ であり、 $Var(SIR) = Var(M \times 100) = 10\,000 Var(M)$

従って、式(11.19)から

$$Var(SIR) = 10\,000 \frac{\sum_{i=1}^A r_i}{\left(\sum_{i=1}^A a_i n_i\right)^2} = \frac{10\,000 \times 16\,404}{(11\,427.97)^2} = 1.2561$$

さらに、 $s.e.(SIR) = \sqrt{Var(SIR)} = 1.12$

従って、この例では、SIRは144であり、その標準誤差は1.12である。SIRの95%信頼区間は、従って、

$SIR \pm (Z_{\alpha/2} \times s.e.(SIR)) = 144 \pm (1.96 \times 1.12)$ 、即ち(141.8, 146.2)

例10. 標準化罹患比の95%信頼区間の近似値の計算

この計算のためのデータは、すでに表9に示した。

観察値 = 16 404

期待値 = 11 427.97

95%信頼区間では、 $Z_{\alpha/2}$ は1.96

下限 = $\left[\sqrt{16\,404} - (1.96 \times 0.5)\right]^2 / 11\,427.97 = 1.41353$

上限 = $\left[\sqrt{16\,404} + (1.96 \times 0.5)\right]^2 / 11\,427.97 = 1.45747$

SIRは、パーセントで表されるので、この近似限界は、141.4と145.7となる。

これらの限界値は、計算がより簡単である利点をもつうえ、答えは、例9で得られた正確な値に極めて近い。

例11. 間接的に標準化した罹患比の有意性の検定

例9では、1980-84年のスコットランドのSIRは144%で、標準誤差は1.12%であった。従って、95%信頼区間は141.8から146.2であり、100を含まない。

同様に、SIRの99%信頼区間は、 $144 \pm (2.58 \times 1.12)$ 、即ち、141.1, 146.9であり、これも、100を含まない。

従って、1980-84年に観察された肺がんの率は、1960-64年より1%の危険率で有意に高いと結論できる。

標準化比と期待値との差の検定

期待値との差の検定は、信頼区間を計算し、その間に100を含むか含まないかをみればよい。

間接法による調整では、標準化の過程で用いた人口の重みは、研究対象集団の年齢階級別人口であるということに留意すべきである。従って、年齢構成の異なる多くのサブグループ（県別、人種別等）のSIRを計算した場合、各々のSIRは例11に示すように、標準人口と比較することはできるが、お互い同士を比較することはできない。スコットランド男性の肺がんのSIRを例とすると、1960-64年を基準として求めた1970-74年のSIRが1.22、1980-84年のSIRが1.44である場合、この値を用いて1970-74年に対する1980-84年のSIRは $1.44/1.22$ 、即ち、1.18であると推測することはできない。

累積率および累積危険度

Day (1987) は、別の年齢調整罹患率として、累積率を提唱した。「5大陸のがん罹患」シリーズ第4巻では、ヨーロッパおよびアフリカ標準人口による計算の代わりに、この指標が用いられている (Waterhouse *et al.*, 1982)。

累積危険度は、他の原因で死亡しない場合に、一定の年齢までに、そのがん罹患する危険率である。危険率が累積される年齢の期間を特定することが必須であり、通常は、0-74歳を用いている。小児がんの場合は、0-14歳を用いる。

累積率は、全年齢階級の年齢階級別罹患率の総和であり、生まれてから74歳までの0-74歳率で計算できる。全年齢階級が同じ人口である場合の直接年齢調整率として解釈できるし、また、累積危険度の近似値とも解釈できる。

i 番目の年齢階級の年齢階級別罹患率は a_i であり、その年齢階級には t_i 年含まれているとする。即ち、0、1-4、5-9...の年齢階級を用いている場合、 t_1 は1、 t_2 は4、 t_3 は5である。累積率は以下の式で表される。

$$\text{Cum.rate} = \sum_{i=1}^A a_i t_i \quad (11.23)$$

年齢階級Aまで合計する。年齢階級別率の計算で、全年齢で5歳階級が用いられているとすると、0-74歳の累積率は $A=15$ となり、

$$\text{Cum.rate (0-74)} = \sum_{i=1}^{15} 5a_i$$

人口10万人あたりで示すよりも、パーセントで表現されることが多い。

Day (1987) によると、累積危険度は次式で表される。

$$\text{Cum.risk} = 100 \times [1 - \exp(-\text{cum.rate}/100)] \quad (11.24)$$

例12. 累積率と累積危険度の計算

表10に、1980-84年のスコットランド男性の肺がんの累積率と累積危険度の計算のためのデータを示す。

表10. 肺がんの累積率と累積危険度（0-74）の計算—スコットランド、男性、1980-84年

年齢階級	年齢階級別率 対10万人	年齢階級の幅	年齢階級別率× 年齢階級の幅 対10万人
	a_i	t_i	$a_i t_i$
0-4	0.00	5	0
5-9	0.00	5	0
10-14	0.00	5	0
15-19	0.00	5	0
20-24	0.37	5	1.85
25-29	0.33	5	1.65
30-34	3.26	5	16.30
35-39	7.48	5	37.40
40-44	21.12	5	105.60
45-49	53.20	5	266.00
50-54	128.16	5	640.80
55-59	263.17	5	1315.85
60-64	422.49	5	2112.45
65-69	600.04	5	3000.20
70-74	780.55	5	3902.75
75-79	—	—	—
80-84	—	—	—
85+	—	—	—
		合計	14 000.85

$$\text{Cum.rate} = \sum_{i=1}^{15} a_i t_i = 11\,400.85$$

累積率（0-74）は10万人あたり11 400.9、即ち11.4%。

累積危険度（0-74）は $100 \times [1 - \exp(-11.4 / 100)] = 10.8\%$

従って、他の死因がなければ、スコットランドの男性は、75歳までに10.8%の危険率で肺がんになる。

累積率の標準誤差

累積率の分散および標準誤差は、直接調整率の分散と標準誤差の式（11.10及び11.11）から、適当な重み（即ち年齢階級の幅、 t_i ）とポアソン近似を用いて、求めることができる。

$$\text{Var}(\text{cum.rate}) = \sum_{i=1}^A (a_i t_i^2 / n_i) \quad (11.25)$$

従って、累積率の標準誤差、s.e.(cum.rate)は下式となる。

$$\text{s.e.}(\text{cum.rate}) = \sqrt{\sum_{i=1}^A (a_i t_i^2 / n_i)} \quad (11.26)$$

累積率の95%信頼区間は式(11.8)を用いて簡単に計算できる。

11.4 ± (1.96 × 0.105)、即ち、11.6, 11.2

例13. 累積率の標準誤差の計算

表11に、累積率の標準誤差を算出するのに必要な計算を示す。

表11. 肺がんの累積率(0-74)の標準誤差の計算—スコットランド、男、1980-84年

年齢階級	年齢階級別率 対人口10万人	年齢階級 の幅	観察人年 n_i	$a_i t_i^2 / n_i$
0-4	0.00	5	827 400	0
5-9	0.00	5	856 500	0
10-14	0.00	5	1 061 500	0
15-19	0.00	5	1 157 400	0
20-24	0.37	5	1 074 900	0.00001
25-29	0.33	5	917 700	0.00001
30-34	3.26	5	890 300	0.00009
35-39	7.48	5	816 000	0.00023
40-44	21.12	5	724 400	0.00073
45-49	53.20	5	706 800	0.00188
50-54	128.16	5	703 800	0.00455
55-59	263.17	5	691 200	0.00952
60-64	422.49	5	610 900	0.01729
65-69	600.04	5	511 800	0.02931
70-74	780.55	5	425 600	0.04584
合計				0.10947

$$\text{var}(\text{cum.rate}) = \sum a_i t_i^2 / n_i = 0.10947 \text{ 対10万人} = 0.00010947\%$$

累積率(0-74)の標準誤差、s.e.(cum.rate)は、この式の平方根によって求める。

$$\text{s.e.}(\text{cum.rate}) = \sqrt{0.00010947} \text{ per } \sqrt{100} = 0.105\%$$

即ち、スコットランド男性の累積率(0-74)は11.4%で、その標準誤差は0.1%であった。

第2部 割合法

パーセント（相対）頻度

登録症例のベースとなる人口がわからない場合は、罹患率を計算することはできない。その場合、異なる症例群間の比較は、がんの種類別の頻度の分布を用いて行わなければならない。よく用いられる方法は、全がんに対する各々のがんのパーセント度数（相対頻度；relative frequency）を計算することである。

$$\text{相対頻度} = \frac{R}{T} \quad (11.27)$$

但し R = 研究群におけるそのがんの症例数

T = 研究群における全がん症例数

相対頻度に代わるものとして、頻度比 (Ratio frequency) (Doll, 1968) がある。頻度比では、各々のがんは、全がんに対してでなく、そのがん以外のすべてのがんに対する割合で表現される。

$$\text{頻度比} = \frac{R}{T - R} \quad (11.28)$$

頻度比は、利点がある場合（例えば、全比較群で大きな割合を占めるがんを処理する場合）もあるが、欠点もある。ここでは詳細にはふれない。

相対頻度の比較は、登録室同士で、あるいは登録室内の地区別、人種別、期間別などで行われる。がんの種類別の相対頻度を用いる際に問題となるのは、相対頻度の比較は、危険率の比較の指標とみなされがちであることで、実際は、危険率の比は罹患率の比でしか測定することができない。2つの百分率の比が相対危険率と同等になるのは、全がんの罹患率が等しいときだけである。

図1に例を示すように、全がんの罹患率がほぼ等しいカリとシンガポール中国人の肝がんの罹患率の比（率比）は6.9である。この2集団の肝がんの相対頻度の比は7.3であり、危険率によく近似している。しかし、シンガポール中国人とダカールの肝がんの率比（相対危険度）は1.0であり、百分率の比は0.3となる。これは、肝がん以外のがんが、ダカールでは少ないために、ダカールの全がんの罹患率（37.0/100000）がシンガポール中国人のそれ（126.2/100000）の29%にすぎないからである。

同じ登録室内で、男女間での相対頻度の比較を行う時にも、同様の問題が生じる。全がんに対する割合をみると、女性のみが発生するがん（乳がん、子宮がん、卵巣がん）の罹患は、男性特有のがん（前立腺、精巣、陰茎）と比較して、はるかに多い。しかし、相対頻度を比

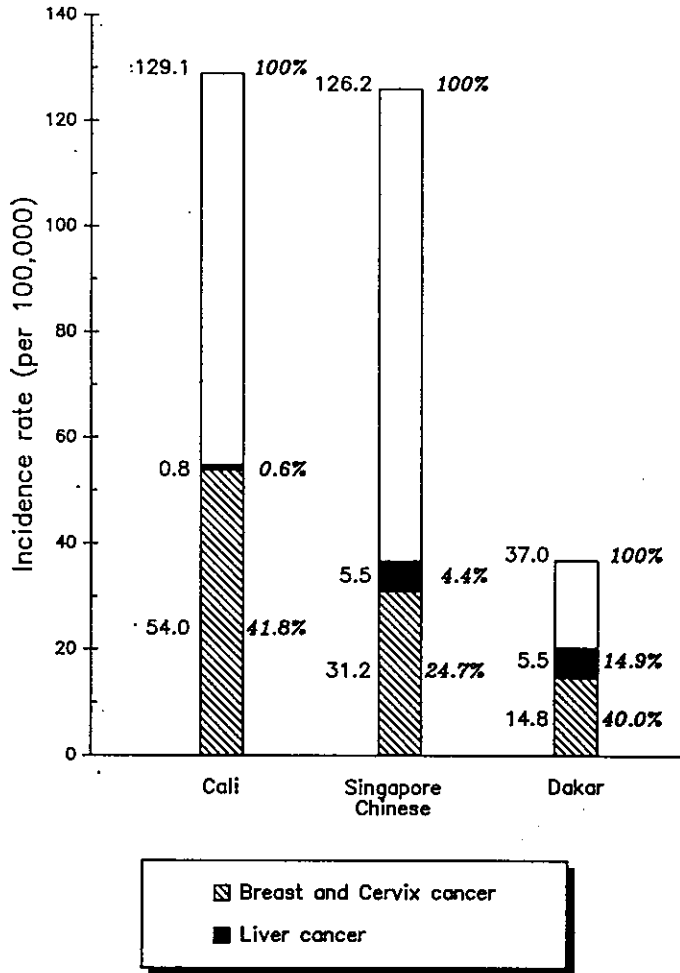


図1. 3ヶ所の地域登録室における女性がんの罹患率（人口10万対）とパーセント分布の比較
 Breast + Cervix cancer (ICD 174+180): liver cancer (ICD 155)
 肝がんの罹患率の比は、シンガポール中国人対カリ = $5.5/0.8 = 6.9$, シンガポール中国人対ダカール = $5.5/5.5 = 1.0$: 百分率比で見ると、シンガポール中国人対カリ = $4.4/0.6 = 7.3$, シンガポール中国人対ダカール = $4.4/14.9 = 0.3$

較する際、全体のパーセントは必ず100になるので、両性で共通のがんの頻度は、女性の方で必ず低くなる。

例えば、カリにおける男性の胃がんの危険率を女性と比較すると、罹患率の比較では1.49となる（図2）。しかし、相対頻度の比は2.03である。なぜなら、女性では性特有のがんが、全がんの約半分を占めているのに対して、男性では12%に過ぎないからである。もちろん、同一性間での比較では、相対頻度の比較は罹患率の比較と同じ結果となる。

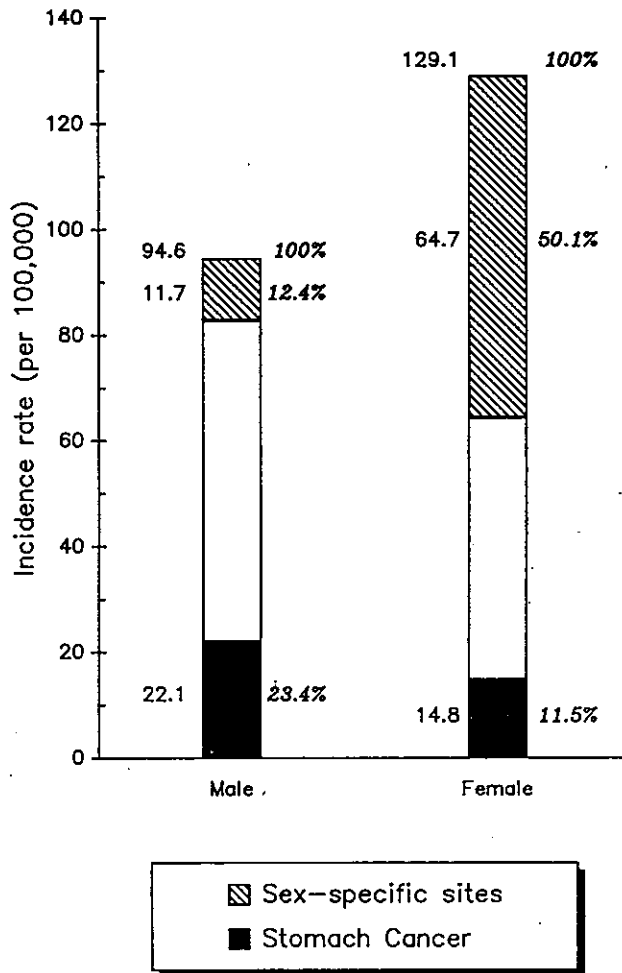


図2. 胃がんおよび性特異のがんの罹患率の男女間の比較—コロンビア，カリ市，1972-76年
 性特異のがんは、女では ICD 174-183，男では 185-187，胃がんは ICD 151。
 胃がんの性比は、罹患率では $22.1/14.8 = 1.49$ ，百分率比では $23.4/11.5 = 2.03$ ，
 しかし性特異がんを除いた場合の百分率比でみると、胃がんの性比は $26.7/23.0 = 1.16$

がんの発生頻度が異なるがん登録室間で相対頻度を比較する際に生じる問題を解決するひとつの方法は、残渣頻度、即ち、最も率の変動の大きいがんを除いた残りに対する頻度を計算することである。この方法は、少数のがんを、全がんの発生率が異なる群で比較する際には有用かもしれない。例えば、Cook & Burkitt (1977) が、アフリカでの比較に利用している。しかし、解釈が幾分、複雑になり、簡単な相対頻度を用いるのに比べて、結論が不明確となる。図2における例では、性特有のがんを分母から引くことにより、全体の罹患率は女

性より男性の方が高くなり、胃がんの残渣頻度の比は1.16となって、実際の相対危険度(1.49)を低く見積ることになる。

図1ですでに例を示したように、子宮がん和乳がんを合わせると、ダカールでは全体の40%を占めるが、シンガポール中国人では24.7%に過ぎない。これらのがんが分母から除かれると、肝がんの残渣頻度はシンガポール中国人で5.8% (4.4 / (100 - 24.7))、ダカールで24.8% (14.9 / (100 - 40.0))となる。これらの残渣頻度を比較することにより得られる相対危険度の推定値は0.23 (5.8 / 24.8)となり、粗百分率からの比較値(0.3)よりも、さらに実際の値(1.0)から離れる。

年齢調整

罹患率を比較する場合と同様に、割合の比較は、比較する集団の年齢構成の差によって複雑になる。

がんの種類別の相対頻度は、年齢によってかなり変動する。例えば、小児では、急性白血病のような悪性腫瘍が多く、成人の悪性腫瘍の大部分を占めるがん(消化器がんや呼吸器がん等)が少ない。従って、全がんに占める特定のがんの割合は、年齢構成によって強い影響を受けるため、それらを比較する場合には何らかの年齢調整が必要となる。

年齢調整のためには、2つの方法が利用されている。年齢標準化がん比(age-standardized cancer ratio; ASCAR)は、直接法による年齢調整と同等であり(Tuyns, 1968)、標準化割合罹患比(standardized proportional incidence ratio; SPIR or PIR)は間接調整法である。これらの中では、PIRにかなりの利点があり、ASCARは、情報源が全く異なるデータを比較する場合で、比較のための明白な基準がない場合のみに有用である。

年齢調整がん比(The age-standardized cancer ratio: ASCAR)

ASCARは直接標準化法であり、比較する群を標準化するための標準年齢構成割合が必要である。標準人口は、任意に選択してよく、比較症例群における全がんの年齢分布に近い標準割合を用いると、粗相対頻度と年齢標準化がん比が比較的近くなる。発展途上国における頻度を比較する際の標準人口割合(Parkin, 1986)を、表12に示す。

ASCARは、以下のように計算できる。

$$\text{ASCAR} = \sum_{i=1}^A (r_i / t_i) w_i \quad (11.29)$$

但し、

r_i = 研究群の年齢階級別の対象がん数

t_i = 研究群の年齢階級別の全がん数

w_i = 年齢階級別標準人口割合

ASCAR は、年齢階級別のがんの頻度が、標準人口における全がんの年齢分布に適応された場合に見られるがんの頻度を意味する。全がんの罹患率の異なるデータ群を比較する際に生じる問題は、粗相対頻度と同様であり、標準化によって補正されないことは強調されるべきである。

ASCARを比較する統計的問題点は、まだ研究されていない。標準誤差の計算式は、まだ存在しない。

表12. 発展途上国のためのがん症例の標準年齢分布

年齢階級	%
0 - 14	5
15 - 24	5
25 - 34	5
35 - 44	10
45 - 54	20
55 - 64	25
65 - 74	20
75 +	10
全年齢	100

出典：Parkin (1986)

例14. 年齢調整がん比の計算

表13に、チュニジア男性の鼻咽頭がんのASCARの計算のためのデータを示す。式(11.29)から、ASCAR = 10.98であり、(式11.27から得られる)粗の相対頻度、 $344/3073 \times 100 = 11.19\%$ と比較できる。

表13. 鼻咽頭がんの年齢調整がん比の計算—チュニジア、男、1976-80年

年齢階級	症例数		全がんに対する 鼻咽頭がんの割合	標準割合 %	期待値 %
	鼻咽頭がん r_i	全がん t_i			
0-14	16	257	0.062	5	0.31
15-24	37	239	0.155	5	0.78
25-34	22	132	0.167	5	0.84
35-44	60	292	0.205	10	2.05
45-54	88	612	0.144	20	2.88
55-64	76	744	0.102	25	2.55
65-74	40	619	0.065	20	1.30
75+	5	178	0.028	10	0.28
全年齢	344	3073	0.112	100	11.00

割合罹患比(The proportional incidence ratio: PIR)

PIR は、標準となる各がんの年齢階級別割合が利用できる場合（間接法による年齢調整法と同様《間接法による年齢調整では、標準となる1群の年齢階級別罹患率が必要である》）に、データ群を比較するために選択できる方法である。がん登録室内でのサブグループ間、例えば、居住地区別、人種別、職業別、などに比較を行いたい場合がよくある。この場合、手に入れやすい標準年齢階級別頻度は、がん登録室全体の各がんの頻度である。（がん登録室のデータには、研究対象としているサブグループが含まれているので、外部標準を用いるのが望ましいが、実際には、大きい割合（30%以上）を占める群がなければ、あまり問題とならない。）

PIR は、研究群におけるそのがんの観測症例数と期待症例数との比である。SIRと同様に、通常、パーセントで表現される。

対象がんの期待症例数は、対象集団の各年齢階級別の全がん数に、そのがんの標準年齢階級別割合を掛け合わせることにより、求めることができる。式で示すと、

$$PIR = (R/E) \times 100 \quad (11.30)$$

$$E = \sum_{i=1}^A t_i (r_i^* / t_i^*) \quad (11.31)$$

但し、

R = 研究群におけるそのがんの症例数

E = 研究群におけるそのがんの期待数

r_i^* = 標準集団における年齢階級*i* のそのがんの症例数

t_i^* = 標準集団における年齢階級*i* の全がん数

t_i = 研究群における年齢階級*i* の全がん数

Breslow and Day (1987) は、PIR の対数の標準誤差を、次式で表現した。

$$s.e.(\log PIR) = \frac{\left[\sum_{i=1}^A r_i (t_i - r_i) / t_i \right]^{1/2}}{R} \quad (11.32)$$

但し、

r_i = 研究群における年齢階級*i* のそのがんの症例数

式(11.32)の保守的な近似式として、以下の式がある。対象とするがんの割合が非常に小さい場合に適応できる。

$$s.e.(\log PIR) = 1 / \sqrt{R} \quad (11.33)$$

表14のデータに対して、式 (11.32)から標準誤差を求めると、

$$\text{s.e.}(\log \text{PIR}) = \frac{\sqrt{325.03}}{545} = 0.033$$

式 (11.33)の近似値は

$$\text{s.e.}(\log \text{PIR}) = \frac{1}{\sqrt{545}} = 0.043$$

Breslow and Day (1987) は、PIR を用いた統計的推計はすべきでないと勧めている。観察された差に関する統計的意味は、信頼区間を用いて評価できる。

例15. 割合罹患比の計算

表14に示されているデータを用いると、タイ国全体の肝がんの年齢階級別割合を標準とした場合の、タイの1地区における肝がんのPIRを計算することができる。

表14. タイの1地区における男性肝がんのPIRの計算 (Srivatanakul et al., 1988)

年齢	タイ国全体			地区4		
	肝がん r_i^*	全がん t_i^*	割合 r_i^*/t_i^*	肝がん r_i	全がん t_i	期待肝がん $t_i(r_i^*/t_i^*)$
0-4	2	210	0.010	0	9	0.090
5-9	1	143	0.007	0	5	0.035
10-14	4	145	0.028	0	4	0.108
15-19	7	230	0.030	1	12	0.360
20-24	23	265	0.087	2	23	2.001
25-29	50	368	0.136	11	37	5.032
30-34	120	492	0.244	22	57	13.908
35-39	169	685	0.247	31	84	20.748
40-44	314	1077	0.292	52	123	35.916
45-49	383	1540	0.249	107	213	53.037
50-54	470	2155	0.218	95	220	47.960
55-59	388	2093	0.185	66	182	33.670
60-64	323	2161	0.149	74	174	26.100
65-69	230	1910	0.120	41	152	18.240
70-74	148	1631	0.091	27	90	8.190
75-79	69	980	0.070	12	35	2.450
80-84	21	426	0.049	4	15	0.735
85+	5	172	0.029	0	8	0.232
	2727	16683		545	1443	268.812

式 (11.31)から、 $E = \sum t_i (r_i^* / t_i^*) = 268.812$

式 (11.30)から、 $\text{PIR} = (R / E) \times 100 = 545/268.812 \times 100 = 203\%$

例15のPIR 2.03に対する95%信頼区間を計算するために、式 (11.32)の標準誤差を用いると、

$$\text{PIR} = 2.03$$

$$\log \text{PIR} = 0.708$$

$$\log \text{PIRの95\%信頼区間} = 0.708 \pm (1.96 \times 0.033)$$

$$= 0.643, 0.773$$

$$\text{PIRの95\%信頼区間} = 1.90, 2.17$$

PIR と SIR との関係

PIRの計算は、研究対象人口に関する情報を含んでいないので、PIRの上昇は、疾患のリスクの上昇を必ずしも意味しない。単に、比較対象としている集団よりも、割合が高いことを意味するだけである。

PIRとSIRの関係は、幾つかの研究グループによって、経験的に研究されている (Decoufle *et al.*, 1980; Kupper *et al.*, 1978; McDowall, 1983; Roman *et al.*, 1984)。

実際、どの研究グループにおいても、以下の関係が認められた。

$$\text{PIR} = \text{SIR} / \text{SIR (all cancers)}$$

SIR-SIR (all cancers)比は、相対SIRと言われる。従って、相対SIRが100よりも大きい場合は、全がんの罹患率に基づく期待値よりも、そのがんの罹患率が多いことが推察される。従って、SIRが100より小さくても、PIRが100以上になることがあり、また逆の場合もある。

表15に、イスラエルがん登録室からの例 (Steintz *et al.*, 1989) を示す。この例では、アジア生まれのユダヤ人男性は、標準人口 (この場合は全てのユダヤ人男性) よりも全がんの罹患が低く、全がんのSIRは77%である。また、アジア生まれのユダヤ人男性では全ユダヤ人男性よりも肺がんのSIRが低い (86%)。しかし、アジア生まれのユダヤ人男性では、全ユダヤ人男性よりも、肺がんは頻度としては重要であるため、PIRは100を越える。

表15. PIRとSIRとの関係。イスラエルのユダヤ人のがん罹患：
全ユダヤ人男性に対するアジア生まれの男性

部位	観察数	SIR (%)	PIR (%)	相対SIR (%)
全がん	6771	77	100	100
食道	114	105	139	136
胃	693	76	100	99
肝	125	110	140	143
肺	1062	86	112	112

(翻訳：味木和喜子、花井 彩)

第 1 2 章 生存率解析

D. M. Parkin and T. Hakulinen

序

地域がん登録室は、特定地域における全がん症例についての情報を集める。従って、そのようなデータから計算された部位別の生存率は、その地域における平均予後を表し、理論的には、少なくともその地域におけるがん対策の効果の客観的指標となる。それとは逆に、院内登録は、一般的にひとつの機関で治療を受けた患者の転帰に関心があり、治療方法別の効果の評価を目指している。

この章では、主として、地域がん登録のデータののための生存率の計算方法について述べる。しかし、解析方法は、院内がん登録のデータでも同じように適用して、あらゆるがん患者群の生存率を記述するために用いることができる。しかし、複数の治療方法の効果を比較、評価するためには、生存率の記述疫学的解析では不十分であり、厳密に管理された臨床試験によってのみ決定することができる。

症例の定義

生存率解析の第 1 段階は、計算を行う患者の性格を明確に定義することである。これらは、通常、以下のように定義されるだろう。

- がんの種類 (部位、組織)
- 診断の時期
- 性別
- がんの進行度

がんの進行度は、一般的には、要約した区分 - 最大 4 群まで - で表示される。臨床的評価 (第 6 章の項目 23) や、外科的病理組織学的評価 (第 6 章の項目 24) に由来する。結果は、年齢階級別、人種別、治療法別、などの形で表示される。

地域がん登録は、生存率解析を登録地区に居住している症例に限るべきである。なぜなら、治療のために転入してきた患者は、平均とは異なった生存状況を示す非典型的なサブグループとなる可能性がある。

解析に含む症例の性質も、同様に、定義すべきである。例えば、診断の根拠が臨床的背景のみの症例を含めるか否かを決定すべきである。死亡診断書のみ (DCO) で登録された患者に関しては、特別の問題が生じる。DCO 症例では、がんの診断日に関する情報が得られない

(DCO 症例では、登録される罹患日 (第 6 章の項目 16) が必然的に死亡日と同じになり、生存期間が 0 となる)。剖検で初めて発見された症例についても同様の問題がある。

Hanai and Fujimoto (1985) は、この問題について以下のように議論している。DCO として登録された症例がある場合、診断時に登録が行われず、治癒したために（もしくは、少なくとも生存しているために）登録データに含まれていない症例が、DCO に相当する程度に存在するだろうと仮定できる。もしこの仮定が正しければ、DCO 症例を生存率の計算に含むことにより、がん登録室のデータベースには致死例が過剰に含まれることになり、真の値より低い生存率を計算することになる。さらに、罹患日（第6章の項目16）と死亡日（項目32）が同じ日であるため、生存期間はゼロとして考慮される。生命表法による実測生存率（後述）を計算する際に、DCO 症例は、生存期間1年未満の患者に含まれるために、1年生存率は人為的に小さくなる。しかし、DCO 症例が、がん登録症例の中で大きな割合を占める場合、地域登録のデータからそれらの症例を除くことは、生存率がもはや地域におけるがんの罹患の平均的予後を反映していないことを意味する。

病悩期間が死亡診断書に記録されている場合、これを用いて診断日（あるいは罹患日）を計算することができる。そのような場合は、DCO 症例は生存率の計算に含まれるべきである。そうでない場合の選択は曖昧である。通常よく用いられている方法は、DCO 症例を除くことであるが、これは、生存率に関して出版されている仕事のほとんどが、DCO 症例の割合が非常に小さい登録室から出ているためであろう。別の解決方法は、2つの生存率、即ち、DCO 症例を含んだ罹患症例の生存率と、DCO 症例を除いた生存率とを報告することである。いずれにしても、DCO 症例の割合は、生存率の報告書に記載すべきである。

開始日の定義

地域がん登録では、開始日（その日から生存率が計算される）は、罹患日（第6章の項目16）である。院内登録では、入院日が用いられる。生存率が治療の転帰を計測するために用いられる場合には、治療開始日が適当であろう。治療の転帰を比較する臨床研究では、無作為に症例を割り付けた日を用いるべきである (Peto et al., 1976, 1977)。

フォローアップ（予後調査）

生存率を計算するためには、登録症例の生死を確認するために、フォローアップ（予後調査）をしなければならない。

受動的フォローアップ

これは、地域の死亡診断書のファイルを用いて行った登録患者の死亡の確認に頼っている。2つのファイル、即ち登録症例の登録ファイルと人口動態統計の死亡票ファイルとの照合が、がん登録室、あるいは地域・国の人口動態統計部門で行われる。照合過程では、患者の同定

指標として、国民背番号（もし利用できるならば）や、氏名、生年月日、住所等のような幾つかの指標の組み合わせを用いる。

受動的フォローアップでは、人口動態統計部門からがん登録室に死亡通知のなかったがん症例は（換言すれば、マッチしなかった全症例は）、生存していると考えられる。従って、受動的フォローアップの結果は、真の生存率を過大に示す可能性がある。誤差の大きさは、照合過程の正確さと、登録がん症例の他地域への転出率との両方に左右される。転出者の記録のファイルを利用できる場合もあり（例えばフィンランド）、そのような患者を調査対象リストから除くこともできる。

能動的フォローアップ

北米の地域がん登録の中には、がん症例を報告する各病院から、フォローアップ情報を入力しているところがある。これらの院内登録は、患者の主治医を通じて、がん登録症例の予後を毎年次々に調査している。この種の調査はメディカル・フォローアップ（経過観察）と呼ばれる。この種の方法を用いると、生存期間と同様に、生存の質をも評価することができる。

ほとんどの地域がん登録室は、個人をフォローするシステムを持っていないが、患者の生存や死亡を間接的に決定するために、他の目的で組織された調査や資料を使うことができる。従って多くの登録室では、住民登録、全英保健医療計画を統括する登録、健康保険や社会保障登録、選挙人名簿、運転免許登録等を利用している。これらの方法は、メディカル・フォローアップからもれた症例の転帰を調べる方法としても用いられている。

能動的フォローアップでは、調査不能で、生死のわからない症例が存在する。生命保険数理法（後述）を用いて生存率を計算する場合に、これらの症例は、最後に生存が確認された日が含まれる期間の丁度半分の期間だけ生存してその地域にいた（従って、危険人口の一部であった）と仮定する。しかし、調査不能（消息不明）例の多くは生存していることが多い（もし死亡しているならば、登録室は死亡診断書によって、それを知るであろう；注）ので、結果は一般的に生存率を下げる方向に偏る、即ち、真の値を低く評価することになる。消息不明患者の数は最小限に留めるべきである。

生存期間

生存率は、測定開始日に生存している症例のうち、一定期間後になお生存している症例の割合で表す。期間は適宜定めればよいが、最も適当な期間は、対象としているがんの予後に

訳者注：死亡診断書が登録室に届かない地域に、患者が転出した場合は、登録室は、患者の死亡を把握できない。

応じて決める。生存率の解釈にあたっては、生存期間開始時の患者数を考慮すべきである。生存期間開始時の生存患者数が10人未満となる期間の患者を対象とした生存率は、多分、印刷すべきでない。

この章に記載されている方法を用いれば、がん患者群の生存状況全体を記述することができる。この方法論を利用する可能性のある人に対し、1時点以上の生存率を計算することを推奨する。5年生存率は、部位別、性別等の患者群の生存率を比較する際の指標として、便宜上よく用いられてきており、しばしば治癒率の指標とみなされている。しかし、多くの部位のがんでは、5年ではこの目的に短すぎるという証拠がある (Hakulinen et al., 1981)。

生存率の計算

以下の節は、アメリカがん合同委員会によって発行されたパンフレット「がん患者の生存率と転帰の報告 1982」(Reporting of Cancer Survival and End Results 1982)を改訂したものである。

がん登録室は、ある日以前の数年間に登録した症例の生存率を計算したいと、通常は希望するであろう。以下の例で、1985年1月1日までの15年間に黒色腫と診断された非常に小さな患者群(50名)を用いて、計算の原則を示す。これら患者の生存率は、1987年末日まで、即ち、1987年12月31日を最終日とした利用可能な予後情報に基づいて評価される。表1に、必要な基本データを示す。

直接法による計算

患者の生存率を要約する最も簡単な方法は、ある一定期間の最後の時点、例えば5年後に生存している患者の割合を計算することである。この目的には、少なくとも5年間、死亡する可能性に曝露された患者のみ、即ち、観察開始時から最終確認日までの間に、少なくとも5年以上経過した患者のみを用いる。この方法は、直接法として知られている。

表1のデータをみると、1987年にフォローできた患者がいるが、フォローできた月は異なっている。1987年にフォローできた患者は、全員1986年12月31日に生存していたことは確かであるが、全員が1987年末日に生存しているか否かは確定できない。従って1986年12月31日を研究の有効最終日とする。すると、1982年1月1日以降に最初の治療を受けた患者は全て、最終日までに、死亡の危険に曝された期間にはなっていないことを意味する。従って、50人中20人(31~50番)は、直接法による計算から除かれなければならない。

表1における「生死状況」の欄で、少なくとも5年間はリスクがあった30名の患者について記入内容を吟味すると、最終確認日に16名の患者が生存しており、14名が1986年12月以前に死亡していた。しかし、この死亡者のうち、1名(2番)は、死亡までに5年間以上生存していた。従って、30名中17名の患者が、それぞれの最初の治療日から5年生存していたこ

表1 黒色腫の患者50名の予後に関するデータ

患者 番号	性別	年齢	診断日 (月/年)	最終接触		診断 からの 生存年	
				日 (月/年)	生死 状況		
1	M	63	10 / 70	10 / 70	D	M	0
2	M	42	7 / 72	1 / 78	D	O	5
3	M	41	3 / 73	4 / 73	D	M	0
4	F	57	6 / 73	7 / 74	D	M	1
5	M	35	9 / 73	10 / 87	A	-	14
6	F	48	10 / 73	8 / 74	D	M	0
7	M	43	4 / 74	2 / 77	D	M	2
8	F	27	1 / 75	1 / 75	D	M	0
9	F	56	12 / 76	10 / 87	A	-	10
10	F	33	1 / 77	11 / 87	A	-	10
11	F	37	4 / 77	4 / 87	A	-	10
12	F	58	9 / 77	8 / 87	A	-	9
13	M	21	2 / 78	5 / 78	D	O	0
14	M	71	2 / 78	11 / 86	A	-	8
15	F	66	6 / 79	8 / 79	D	M	0
16	F	35	7 / 79	12 / 87	A	-	8
17	F	31	10 / 79	11 / 87	A	-	8
18	M	35	3 / 80	6 / 87	A	-	7
19	F	44	4 / 80	7 / 87	A	-	7
20	M	26	4 / 80	10 / 87	A	-	7
21	M	57	10 / 80	6 / 81	D	M	0
22	M	54	12 / 80	2 / 81	D	M	0
23	M	63	1 / 81	1 / 82	D	M	1
24	F	32	1 / 81	10 / 83	D	M	2
25	F	43	4 / 81	2 / 87	A	-	5
26	F	76	7 / 81	2 / 86	D	M	4
27	M	31	9 / 81	11 / 87	A	-	6
28	M	77	11 / 81	2 / 87	A	-	5
29	F	59	11 / 81	4 / 87	A	-	5
30	F	76	12 / 81	9 / 87	A	-	5
31	M	39	3 / 82	8 / 85	D	M	3
32	F	50	7 / 82	4 / 87	A	-	4
33	F	38	10 / 82	6 / 87	D	M	4
34	F	82	3 / 83	12 / 87	A	-	4
35	M	65	4 / 83	7 / 83	D	M	0
36	M	40	4 / 83	10 / 87	A	-	4
37	M	22	6 / 83	2 / 87	A	-	3
38	F	25	1 / 84	11 / 87	A	-	3
39	M	33	4 / 84	11 / 87	A	-	3
40	F	51	5 / 84	7 / 87	A	-	3
41	F	40	7 / 84	11 / 87	A	-	3
42	M	70	9 / 84	9 / 85	D	O	1
43	M	47	9 / 84	12 / 85	D	M	1
44	M	67	10 / 84	4 / 86	D	O	1
45	F	58	1 / 85	8 / 87	A	-	2
46	M	75	1 / 85	10 / 87	A	-	2
47	M	40	4 / 85	7 / 87	A	-	2
48	F	35	4 / 85	7 / 87	A	-	2
49	F	49	5 / 85	12 / 86	D	M	1
50	F	21	6 / 85	3 / 87	A	-	1

注：(1) 生死：Aは生存、Dは死亡。(2) 死因：Mは黒色腫、Oはその他。

とになり、5年生存率は57%となる。

生命保険数理法による計算

生存率計算の直接法では、利用できる情報の全部を利用していない。例えば、31番の患者は治療開始から4年目に死亡しており、32番の患者は4年以上生存している。そのような情報は有益であるが、直接法の規則に従うと、それらの患者は1981年12月以降に診断されているために利用することができない。

生命保険数理法、あるいは生命表法は、研究の最終日までに累積されたフォローアップ情報を、全て用いることのできる方法である。生命保険数理法は、生存パターン、即ち、全観察期間中に、どのように患者群が減っていくかについての情報を提供できる利点もある。

生命保険数理法は、少ない患者群（この例では50名）の生存状況を注意深く解析したい研究者のために、作られたものである。しかし、多数の症例をコンピュータで解析する場合にも、同じ基本的方法論が用いられている（例えば、Hakulinen & Abeywickrama, 1985）。

実測生存率 (Observed survival rate)

表2に、研究対象の50名について利用できる予後情報の全てを用いた、生命表法による生存率の計算法を例示した。表2のような表を作成するためには、6段階必要である。

- (1) 診断から各期間ごと（表2では1年単位）の患者の生存状況（生存もしくは死亡）と観察途中の者が、第3、第4欄の計算に用いられている。第3欄と第4欄とに入ってくる患者の和は、患者の総数と等しくなければならない。第2欄の、最後の期間（診断から5年以後）の当初に生存している患者17名が、第4欄にも（この期間中の最終生存確認患者数）に書かれている。
- (2) 第2欄には、各期間の当初に生存している患者数を書き、これは連続的に減算することで求められる。つまり、診断された50名の患者の中で、最初の1年間に9名が死亡したので、診断1年後に41名が生存している。第2期間（診断から1年以上2年未満）では、6名が死亡し、1名が追跡不能であるため、3番目の期間（2-3年）の当初には34名の患者が残っている。
- (3) 死亡のリスクをもつ有効生存者数（第5欄）は、それぞれの期間中に最後に生存が確認された患者は、平均するとその期間の中央の時点まで生存が確認された、という仮定に基づいている。従って、第3期間（2-3年）の有効生存者数は、 $34 - (1/2 \times 4) = 32.0$ となり、第4期間（3-4年）では $28 - (1/2 \times 5) = 25.5$ となる。
- (4) 各期間中の死亡割合（第6欄）は、第3欄を第5欄で割ることで求める。従って、

表2. 生命保険数理法（生命表法）による実測生存率およびその標準誤差の計算

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
診断から の期間 (年)	期間当初 の生存数	期間中の 死亡数	期間中の 観察中途 数 —消息 不明—	死亡割合 計算の対 象数 (有効生 存者数)	期間中の 死亡割合	期間末の 生存割合	治療日か ら計算し た当該期 間末での 生存割合	5欄—3欄	6欄/9欄
(i)	(注1) (I _i)	(d _i)	(w _i)	(r _i)	(注2) (q _i)	(注2) (p _i)	(注2) (Pp _i)	(r _i -d _i)	(q _i) / (r _i -d _i)
0-1	50	9	0	50.0	0.180	0.820	0.820	41.0	0.0044
1-2	41	6	1	40.5	0.148	0.852	0.699	34.5	0.0043
2-3	34	2	4	32.0	0.063	0.937	0.655	30.0	0.0021
3-4	28	1	5	25.5	0.039	0.961	0.629	24.5	0.0016
4-5	22	2	3	20.5	0.098	0.902	0.567	18.5	0.0053
5-	17	-	17	-	-	-	-	-	-
計		20	30						0.0177

注1：0-1 は、0年以上1年未満を意味する。以下同様。

注2： $r_i = I_i - w_i / 2$ ； $q_i = d_i / r_i$ ； $p_i = 1 - q_i$

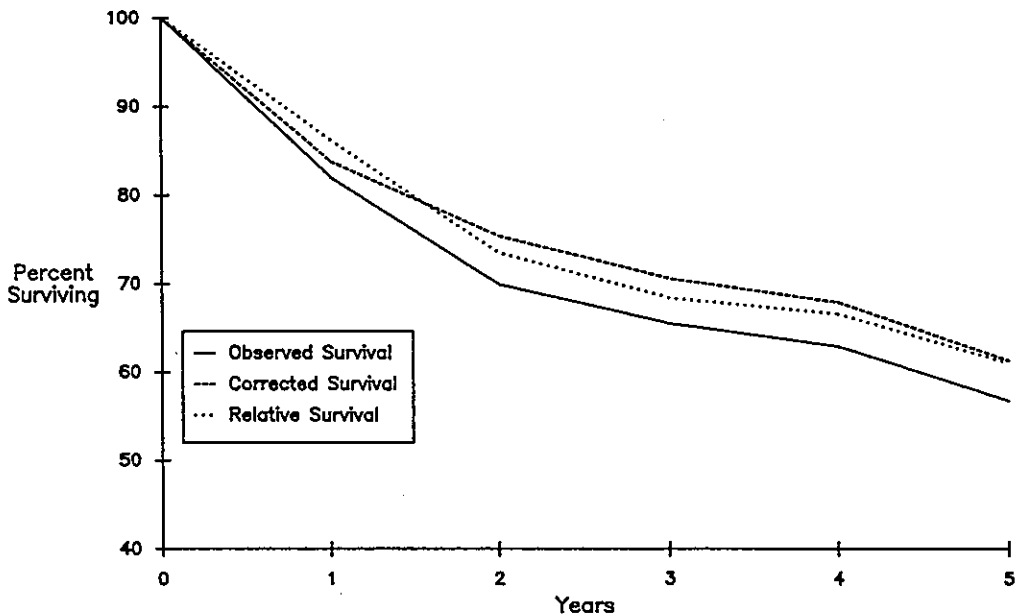


図1. 黒色腫患者の生存率曲線の種類別比較—実測率、補正率、および相対生存率—
(表1, 2のデータより)

診断後0-1年の死亡割合は $9/50.0 = 0.180$ であり、1-2年は $6/40.5 = 0.148$ である。

- (5) 各期間における期間末の生存割合（第7欄）、即ち、実測期間生存率は、1.000から死亡割合（第6欄）を引くことで計算できる。
- (6) 診断から各期間末までの生存割合、即ち、実測生存率（第8欄）は、各生存率とそれ以前の全期間生存率とを掛けたものである。例えば、5年生存率 0.567は、第7欄の数を0-1年から4-5年まで全部掛けたものである。

生命表法を用いて計算した5年生存率は 0.567、57%である。この例では、50名の患者全員の情報を用いて計算した結果と、直接法による30名の患者に基づく結果とは相似している。直接法の計算で除かなければいけない患者がある場合、両者の方法の結果は通常一致しない。その場合、生命表法による結果の方が、より多くの情報に基づいているために、信頼性が高い。

生命表法のひとつの利点は、連続する観察期間における死亡のリスクの変化についての情報を提供できることである。つまり、第6欄 (q_i) をみると、診断から4年目までの各年死亡割合は、第1期間 (0-1年) の18%から第4期間 (3-4年) の4%に減少していることがわかる。(第5期間 (4-5年) に10%に増加しているのは、含まれている人数が少ない、即ちこの期間の当初に生存している患者は22名しかないので、偶然の結果かもしれない。)

第8欄の累積生存率を用いると、生存曲線が描け、生存パターンを図で表現することができる (図1)。

補正生存率 (Corrected survival rate)

前述した実測生存率は、死因に関係なく、全ての死亡を包括している。これは、患者群における全ての死亡の真実の反映であるが、通常は、研究している疾患に起因する死亡を記述することが主目的である。表1をよく見ると、4例では黒色腫が死因でない (2番、13番、42番、および44番の患者)。その内3例は、最初の5年以内にその死亡が起こっているので、表2で計算した5年生存率に影響している。

死因について信頼性のある情報が利用できる場合はいつでも、研究対象疾患以外による死亡を補正することができる。手順を表3に示す。死亡は対象疾患による死亡 (第3欄 a) と、それ以外の死因による死亡 (第3欄 b) とに分けられる。他の死因で死亡した患者は、その期間内に最後の生存が確認された患者 (第4欄) と同等の扱いを受ける。即ち、両者はともに、黒色腫で死亡するリスクから外れる群となる。従って、第2期間 (1-2年) の (黒色腫による) 死亡のリスクとなる有効生存者数 (第5欄) は $41 - (2+1) / 2 = 39.5$ となる。

実測生存率57%に対して、補正5年生存率は 0.613、61%となる。補正率は、黒色腫患者

表3. 補正生存率の計算

(1)	(2)	(3)		(4)	(5)	(6)	(7)	(8)
診断から の期間 (年)	期間当初 の生存数	期間中の死亡数		期間中の 観察中途 数	死亡割合 計算対象 数	期間中の 死亡割合	期間末の 生存割合	治療日か ら計算し た当該期 間末での 生存割合
(i)	(注1) (l _i)	(d(m) _i)	(d(o) _i)	(w _i)	(r _i)	(注2) (q _i)	(注2) (p _i)	(注2) (Pp _i)
0-1	50	8	1	0	49.5	0.162	0.838	0.838
1-2	41	4	2	1	39.5	0.101	0.899	0.754
2-3	34	2	0	4	32.0	0.063	0.937	0.706
3-4	28	1	0	5	25.5	0.039	0.961	0.679
4-5	22	2	0	3	20.5	0.098	0.902	0.613
5-	17	-	-	17				
計		17	3	30				

第5-8欄の '死亡' および '生存' は、対象としている疾患による死亡 (および生存) を意味する
注1: 0-1 は、0年以上1年未満を意味する。以下同様。

注2: $r_i = l_i - (w_i + d(o)_i) / 2$; $q_i = d(m)_i / r_i$; $p_i = 1 - q_i$

の61%が、診断から5年以内に黒色腫で死亡するリスクから逃れたことを意味する。

補正率を用いることは、対象疾患以外の死因で死亡する可能性に強く影響する可能性のある要因、例えば、性、年齢、人種、社会経済状態等、の異なる患者群を比較する際に、特に重要となる。図1で、50名の患者の実測生存率と補正生存率とを比較しているが、実測および補正生存率曲線の間のギャップは、通常 (黒色腫以外) の死亡を表している。

相対生存率 (Relative survival rate)

死因に関する情報が利用できなかつたり、信頼性が低い場合がある。その場合、補正生存率を計算することができない。しかし、通常の死亡の期待値についての患者群間の差、即ち、研究対象疾患以外の死因で死亡するリスクの差を計算することは可能である。これは、相対生存率を利用することで達成できる。この相対生存率は、人種、性、年齢、観察年について、患者群と似た構成をもつ一般住民群で求めた期待生存率に対する実測生存率の比である。

期待生存確率は、出版されている一般住民の生命表から、毎年の生存確率を掛け合わせることで得ることができる。個々の患者の性、年齢と、登録年とに応じて、それぞれの患者に適合した確率を生命表から得ることができる。表4に、1番目の患者 (1970年に63歳の男性) の期待5年生存確率を計算するのに必要なデータ (フィンランドの成績) を示す。フィンランドでは、5年毎に1歳階級別の一般人口の年間死亡率が刊行されており、5年間の平均を示している。1番目の患者は、1966-1970年の期間 (実際は1970年) では63歳であり、その後5年間 (1971-1975年の生命表でカバーされる) 生存した。出版された一般人口の生命表 (Official Statistics of Finland, 1974, 1980) から、患者が年をとるごとに、それに対応した一般人口の死亡率を得る。これを患者の通常の死亡確率とみなす。対応する通常の生存確率を計算するために、これらを1から減ずる。患者は、観察開始時には、丁度63歳ではなく、

表4. 一般人口の死亡率を用いた期待5年生存確率の計算（フィンランド）

年齢	歴年	年間死亡率(注)	年間生存確率	2年移動平均
63	1966-70	36.08	0.96392	
64	1971-75	35.28	0.96472	0.964320
65	1971-75	38.75	0.96125	0.962985
66	1971-75	41.40	0.95860	0.959925
67	1971-75	46.24	0.95376	0.956180
68	1971-75	48.52	0.95148	0.952620

注：人口1000に対する年間死亡確率 (Official Statistics of Finland, 1974, 1980)

平均して63.5歳に近いという事実を酌量して、毎年の通常生存確率から移動平均を計算する。63.5歳から67.5歳に対応する5つの生存確率を相乗して、5年期待生存確率を求める。この例では結果は0.812となる。

表1の患者群全体では、平均期待生存率は、各々の患者の5年生存確率を総和し、それを患者数で割ったものである。これが0.94、94%であるとする、

$$\text{相対生存率} = \frac{\text{実測生存率}}{\text{期待生存率}} \times 100 = 0.57 / 0.94 = 61\%$$

この例では、相対生存率は補正生存率と同じである。

実際は、各期間の相対生存確率を計算して、連続するフォローアップ期間で累積する方法が一般的である (Ederer ら、1961参照)。

相対生存率を利用する場合は、死因に関する正確な情報(がんが死因であるか、偶然に他の死因で死亡したのか)を必要としない。このことは、大きな利点となりうる (Hakulinen, 1977)。しかし、調査対象集団が生命表の計算に用いられている一般住民と同じ死亡リスクにさらされていることが前提条件となる。適当な生命表が利用できない場合(例えば、ある人種とか社会経済群など)は、補正生存率が望ましい。とにかく、どの方法によったかは明示すべきであり、異なる患者群間で生存率を比較する場合は、すべての患者群に同じ方法を用いるべきである。

10年以上フォローアップした場合の生存率の計算に相対生存率を用いる場合は、Hakulinen (1982) の論文を参照すべきである。この論文では、基本とした集団の老化による偏りと、年齢階級別罹患率推移の差による偏りの、補正の仕方について述べている。

表5. Kaplan-Meier 法による実測生存率の計算

診断から の月数 (l)	月初の 生存数 (l_i)	期間中の 死亡数 (d_i)	観察中途数 (w_i)	期間中の 死亡割合 (q_i)	期間末の 生存割合 (p_i)	累積生存率 (P_i)
0	50	2	0	0.040	0.960	0.960
1	48	1	0	0.021	0.979	0.940
2	47	2	0	0.043	0.957	0.900
3	45	2 *	0	0.044	0.956	0.860
8	43	1	0	0.023	0.977	0.840
10	42	1	0	0.024	0.976	0.820
12	41	2 *	0	0.049	0.951	0.780
13	39	1	0	0.026	0.974	0.460
15	38	1	0	0.026	0.974	0.740
18	37	1 *	0	0.027	0.973	0.720
19	36	1	0	0.028	0.972	0.700
21	35	0	1			
27	34	0	2			
30	32	0	1			
33	31	1	1	0.032	0.968	0.677
34	29	1	0	0.034	0.966	0.654
38	28	0	1			
40	27	0	1			
41	26	1	0	0.038	0.962	0.628
43	25	0	1			
44	24	0	1			
46	23	0	1			
54	22	0	1			
55	21	1	0	0.048	0.952	0.598
56	20	1	0	0.050	0.950	0.568
57	19	0	2			
≥ 60	17	1 *	16			

*黒色腫以外の死亡1例

Kaplan-Meier 法による計算

生存率の計算に広く利用されており、多くのコンピュータプログラムで利用できる方法として、Kaplan-Meier 法がある(Kaplan & Meier, 1958)。この方法は、生命保険数理法に似ているが、フォローアップ中の各年の累積生存率のかわりに、死亡日の記録の精密さに応じて、短い期間での生存患者の割合を計算することができる。

表1のデータを用いて、方法を表5に示す。各々の死亡例(d_i)と、調査不能、打ち切り例(w_i)とは、これらが発生した日に最も近い区分(月)で計算に加えられている。計算法はほとんど生命保険数理法と同じであるが、期間が1月単位となり、また調査不能、打ち切り例は最終確認月の最後まで生存していたと仮定される点異なる。

Kaplan-Meier 法で計算された生存曲線を図2に示す。各々の死亡に対応する縦線と横線の階段から構成され、生命保険数理法による線グラフとは対照的である。

最初の5年間に発生した3人の黒色腫以外の死亡を打ち切り例(w_i)として扱うことにより、この方法で補正率を計算することができる。相対生存率は、生命保険数理法の場合と同様に、実測生存率を期待生存率で除することにより計算することができる。

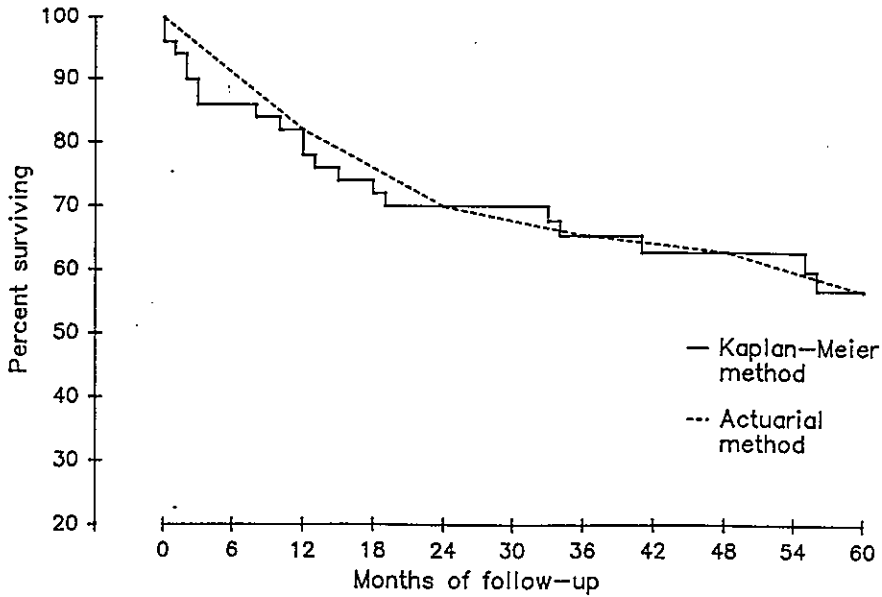


図2. 黒色腫患者の Kaplan-Meier 法での生存率曲線
(生命表方式により計算した実測生存率との比較)

生存率の年齢補正

補正生存率や相対生存率を利用することにより、部分的には年齢補正が達成されている。なぜなら、それらの方法は、年齢とがん以外の死因での死亡との関係を酌量しているからである。しかし、年齢と観察対象のがんによる死亡リスクとの間に何らかの関連があり、年齢構成の異なる集団間での比較を行いたい場合は、罹患率の場合と同様に、年齢階級別生存率の比較に限るか、年齢調整の手法を用いるべきである (Haenszel, 1964)。

生存率の標準誤差

罹患率の場合 (第11章) と同様に、生存率の正確性の指標として、標準誤差と信頼区間とが用いられている。

直接法による生存率の標準誤差の計算

$$s. e. (P) = \sqrt{[P(1-p) / N]} \quad \text{但し、} P = \text{生存率、} N = \text{対象者数}$$

直接法による生存率の計算では、5年間観察した患者数は30であるので

$$s. e. (P) = \sqrt{[0.57 \times (1-0.57) / 30]} = 0.090$$

95%信頼区間は、以下の通りである。

$$P \pm 1.96 \times s. e. (P) = 0.57 \pm 1.96 \times 0.09 = 0.39 \sim 0.75$$

生命保険数理法による標準誤差

生命保険数理法による5年生存率の標準誤差の計算には、表2の最後の2欄の値を用いる。第9欄は、第5欄から第3欄を減ずることで求められ、第10欄は、第6欄を第9欄で除することで求められる。第10欄の総和は0.0177となる。生命保険数理法による5年生存率の標準誤差は、5年生存率に第10欄の総和の平方根を乗ずる、即ち、 $0.567 \times 0.0177 = 0.075$ となる。表2の略語を用いて式で示すと、

$$s. e. (P) = P \sqrt{(\sum [q_i / (r_i - d_i)])} = P \sqrt{(\sum [d_i / r_i (r_i - d_i)])}$$

この式はGreenwoodの式として知られている。

従って、患者の5年生存率の95%信頼区間は、 $0.567 \pm 1.96 \times 0.075$ 、即ち、0.42～0.72となる。

実際は、生命保険数理法の標準誤差の近似値は、Ederer (1960) が作成した表から簡単に得ることができる。

生命保険数理法から得た5年生存率の標準誤差は、直接法による生存率の標準誤差より小さくなることに留意すべきである(0.075対0.090)。この差は、利用できる全ての情報、即ち、観察期間5年未満の患者の情報をも用いることにより生じる統計的緻密性、という意味での生命表法の利点を反映している。

さらに詳しい情報についてはMerrell & Shulman (1955) や Culter & Ederer (1958) を参照のこと。

相対生存率の標準誤差

相対生存率の標準誤差は、実測生存率の標準誤差(直接法あるいは生命保険数理法で得たもの)を、期待生存率で割ることにより、簡単に求めることができる。生命保険数理法による5年実測生存率が57%、期待生存率が94%の場合、相対生存率は61%となる。実測生存率の標準誤差は0.075である。

この例の場合、5年相対生存率の標準誤差は

$$\frac{\text{実測生存率の標準誤差}}{\text{期待生存率}} = \frac{0.075}{0.940} = 0.080$$

5年相対生存率の95%信頼区間は、従って、 $0.61 \pm 1.96 \times 0.080 = 0.45 \sim 0.77$ 。

生存率の比較

今、最も単純な条件下で2つの生存率を比較する場合、2つの生存率の95%信頼区間が重なっていないければ、観察された差は慣習的に統計的に有意である、即ち、偶然の結果ではないと考える。しかし、この方法はあまり推薦できない。より適切な方法を以下に示す。

標準的な統計学のテキストは z 検定について述べている。z 検定は、もし偶然のみが作用しているならば、2 者の差が観察値と同じ、あるいは観察値以上の値をとる確率の推定値を提供する。z 値は次式で計算できる。

$$z = \frac{|P_1 - P_2|}{\sqrt{[(s.e.(P_1))^2 + (s.e.(P_2))^2]}}$$

但し、

P_1 = グループ 1 の生存率

P_2 = グループ 2 の生存率

$|P_1 - P_2|$ = 差の絶対値、即ち、正負にかかわらない差の大きさ

s. e. (P_1) = P_1 の標準誤差

s. e. (P_2) = P_2 の標準誤差

z 値は標準正規分布に従うので、もし z が 1.96 よりも大きければ、観察されたのと同様の差が偶然に起こる確率は 5% 未満であり、z が 2.56 より大きければ、その確率は 1% 未満となる。

例えば、表 6 に、24 名の男性 ($P_1=0.485$) と 26 名の女性 ($P_2=0.646$) の、生命保険数理法による 5 年実測生存率の計算を示す。Greenwood の式を用いると、 P_1 の標準誤差は 0.105、 P_2 の標準誤差は 0.105 となる。従って、

$$z = \frac{|0.485 - 0.646|}{\sqrt{(0.105^2 + 0.105^2)}} = \frac{0.161}{0.148} = 1.09$$

求められた z 値は 1.96 よりも小さいので、5% レベルで統計的有意差がない。この程度の生存率の差が統計的に有意になるためには、対応する標準誤差が小さくなるように、より多くの患者を研究に含める必要がある。

いくつかのグループ間の生存率を比較するさらに良い検定方法は、logrank test である (Peto et al., 1977; Breslow, 1979 参照)。この検定法は、フォローアップの一時点における比較(前述の例のように)に限定されず、フォローアップ全期間の情報を利用する。この方法は、臨床研究での治療方法別の生存状況の比較によく用いられる。通常、Kaplan-Meier 型の生存率曲線 (図 2) を描くことができるように、各患者の診断から死亡までの期間が正確に知られている。しかし、例示するために、2 群 (男性および女性) の毎年の生存率を示す表 1 のデータに、logrank test の近似法を適用することが出来る。この近似法は控えめな方法であるので、必ずしも適度に小さな p 値を得ることができるとは限らない (Crowley & Breslow, 1975)。ほとんどすべてのプログラムパッケージには logrank test が含まれているので、その利用を勧める。

各期間での期待死亡数を、各群で計算する。この計算には、各群の有効生存数 (r_i) と、全群をまとめたその年の死亡確率 (Q_i) [表 6 では男性と女性を合わせたもの (表 2 の第 6

表6. 実測生存率および年間の期待死亡数の計算, 男女別

年	l_i	d_i	w_i	r_i	q_i	p_i	Pp_i	期待死亡数 ($r_i \times Q_i$)
男								
0	24	6	0	24.0	0.250	0.750	0.750	4.32
1	18	4	0	18.0	0.222	0.778	0.584	2.66
2	14	1	2	13.0	0.077	0.923	0.539	0.82
3	11	1	2	10.0	0.100	0.900	0.485	0.39
4	8	0	1	7.5	0.000	1.000	0.485	0.74
5	7	1	1	6.5	0.154			0.43
		13						9.36
女								
0	26	3	0	26.0	0.115	0.885	0.885	4.68
1	23	2	1	22.5	0.089	0.911	0.806	3.33
2	20	1	2	19.0	0.053	0.947	0.764	1.20
3	17	0	3	15.5	0.000	1.000	0.764	0.60
4	14	2	2	13.0	0.154	0.846	0.646	1.27
5	10	0	3	8.5	0.000			0.57
		8						11.65

注: Q_i は年間に死亡した全例 (男および女) の割合 (表2の第6欄)

欄) を利用する。各群の期待死亡数は、各期間の期待死亡数の総和となる。

$$\text{期待死亡数} = \sum r_i Q_i$$

生存率曲線の同一性の検定は、カイ2乗検定を用いておこなうことができる。サブグループが j 群ある場合、自由度は $(j-1)$ となる。

$$\chi^2 = \sum \frac{|O_i - E_i|^2}{E_i}$$

例えば、表6の男性と女性とを比較する場合、全ての死亡についての情報を利用する (これらは全て6年未満の期間に含まれる)。

男性では、5年後の観察死亡数=13
5年後の期待死亡数=9.36

女性では、5年後の観察死亡数=8
5年後の期待死亡数=11.65

$$\chi^2 = \frac{|13 - 9.36|^2}{9.36} + \frac{|8 - 11.65|^2}{11.65} = 2.56$$

自由度1では、 $p > 0.1$ であり、有意差のない結果である。

相対生存率曲線のための検定方法が、Brown (1983)と Hakulinen et al. (1987) によって作られている。これらは、Hakulinen and Abeywickrama (1985) のコンピュータソフトウェアで利用することができる。

患者群間の生存率を比較する場合、どの疫学的研究でも必要であるのと同様に、交絡因子のコントロールが必要である場合が多い。例えば、ある病院で治療を受けた患者の生存率と、別の病院のそれとを比較する場合に、予後に影響する可能性のある要因 (例えば、年齢、人

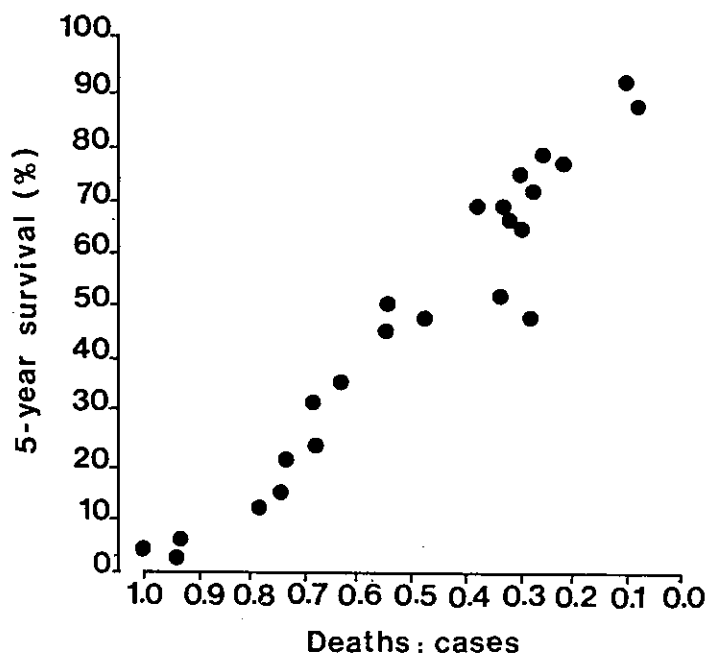


図3. 5年相対生存率（1973-76年登録患者）と死亡者の割合（1973-77年）との相関：24部位、SEERのデータ

種、社会的地位、がんの進行度等）の2群における差を考慮する必要がある。これを行うひとつの方法は、交絡因子別に層化することであるが(Mantel, 1966)、最近は、比例ハザードモデル (Cox, 1972)を基本としたモデル化技術の利用が増えている。このモデルのコンピュータプログラムは、全ての主要な統計パッケージに含まれている。この方法の相対生存率への一般化は、Pocock et al. (1982) や Hakulinen and Tenkanen (1987)が行っている。後者は、GLIM(Baker & Nelder, 1978)に基づき、非比例ハザードにも順応している。

致命率

がん登録室の中では、生存率を計算するための登録症例の総合的なフォローアップが不可能な登録室も多くある。しかし、そのような場合にも、生存率の指標として致命率、即ち、一定期間内に診断された症例に対する同期間の死亡症例の割合を表示することができるかもしれない。同じ率が、「期間内の死亡」(Muir & Waterhouse, 1976)、あるいは近年は「死亡/罹患比」(Muir et al., 1987)として、「五大陸のがん罹患」シリーズの中で、登録の完全性の指標として用いられている。もちろん、罹患症例と死亡症例とは、必ずしも同一の症例ではなく、単に同じ診断であるというだけであり、この比は一般の生存経験の間接的な指標に過ぎない。しかし、図3に示すように、同一登録室内の各がん別の5年生存率と致命率との関係は、実際、非常によく似ることが多い。しかし、異なる登録室間での生存率の比較に致命率を用いることができるか否かは不明確である。

謝辞

著者らは、米国国立がん研究所、Division of Cancer Prevention and Control, Biometry Branch の Dr. M. Myers から、この章の用意に際して、役に立ったコメントと提言を戴いた。感謝する。

(翻訳：味木和喜子、花井 彩)

第 1 3 章 院内がん登録

J. L. Young

序

院内がん登録の目的は、地域がん登録の目的とは、当然異なる。研究及び企画立案における地域がん登録の利用については、第3章で述べた。院内がん登録の目的は、病院管理、ならびに、病院固有の対がん医療計画 (cancer program) の面からの要請、そしてとりわけ個々の患者からのニーズ、などに対応することである。各院内がん登録の設立は、歴史的には、登録の存在によって個々の患者への診療が向上するという信念に由来している。なぜなら、登録が、フォローアップ検査のために患者が定期的に受診することを確実にするのに役立つからである。実際、世界の院内がん登録の中には、フォローアップの予約表を作成することが腫瘍登録士の責任となっているところもある。

以上述べたように、院内がん登録は、病院管理とがん患者の医療とを指向している。それ故、院内がん登録で収集されたデータ項目の中には、地域がん登録で収集されるものと異なる項目がある。逆にいえば、多くの院内登録は、それぞれのデータを地域がん登録にも提出するので、院内がん登録にとっては必要のないものでも、地域がん登録が必要としているデータ項目を、しばしば含めなければならない。こうした状況を詳しく述べる。

病院内では、院内登録が、院内対がん医療計画やがん予防、医療サービス供給システムにとって、しばしば欠くことのできないものとされている。米国では、例えば、アメリカ外科学会が、効果的な認可機構を持ち、それにより個々の第一線病院の対がん計画を認定している。米国内の1,200以上の病院が、この認定を得ている。いかなる病院対がん医療計画であれ、アメリカ外科学会から認定を受けるためには、次の4つの主要構成要素を持っていないといけない。すなわち、① 病院がん委員会、② 定期的に行われるがんカンファランス、③ 患者ケアの評価研究、④ がん登録、である。

このような枠組みの中で、がん登録は、様々な機能への積極的な参加を通して他の3つのプログラムに便宜を与え、そしてがん委員会に対して直接的に責任を負っている。がん委員会は、病院内の常任委員会で、その構成は学際的でなければならず、かつその義務と責任を明確にしていなければならない。それ故、院内登録は、がん委員会がその義務と責任（その責任は、定期的な教育カンファランスを組織、提供、指導、評価することから、がん予防プログラムのニーズを決定し、患者に対して直接的なサービスを提供しつつ、がん医療の評価研究にまでわたる）を遂行することを支援するように、組織されている。

院内登録の機能の一つは、病院理事会に対する年間の対がん活動の年次報告を作成し、病

院のがん診療実態に関する事項を記録することである。アメリカ外科学会は、認定プログラムにおいて、それぞれの病院のデータを全国データと比較するように求めており、そうすることで外科学会は、その病院のがん診療実態が一般人口での実態と比較してどうであるかを把握できる。こうした年次報告を作成するにあたり、正確にどのがん症例を含めるのかを熟慮しておかなくてはならない。その年を通じて、あらゆる時期に診療したがんのすべての患者を含めるべきか？相談だけの目的で診療した患者を含めるべきか？他の病院で、既に診断され、もしくは、治療を受けている患者を含めるべきか？こうした点の考慮は、「症例区分」という概念の由来になっている。「症例区分」は、院内登録にとっては非常に重要なデータ項目の一つであるが、地域がん登録にとっては、全患者についての地域ベースのフォローアップを実施しないのであれば、意味がない。地域ベースのフォローアップにおいては、個々の病院がフォローに責任のある患者を示すのに、「症例区分」を使うことが出来る。一般的に受け入れられている症例の6区分を、以下に示す。

- (1) 院内登録の責任（開始）日以降、当該病院で診断されたが、初回治療のすべてが他の医療機関で行われた症例。
- (2) 当該病院で診断と治療が行われた症例（注意：患者が治療不能とみなされても、この範疇に含まれる）。
- (3) 他の医療機関で診断されたが、初回治療のすべて、あるいは一部が当該病院で行われた症例。
- (4) 他の医療機関で、診断と初回治療のすべてが行われた症例（支持療法のための目的で入院した患者を含む）。
- (5) 院内登録の責任（開始）日以前に、当該病院で診断と治療を受けた症例。
- (6) 剖検でのみ診断された症例。

(1)、(2)、(3)の範疇に入る症例は、一般に解析症例と呼ばれ、これらは、がん患者の医療の点において、病院がいかにか成功しているかを評価するための定期報告の作成に含められる。(4)から(6)までの症例は、非解析症例として扱われ、ほとんどの表、特に生存率の算出から除外される。しかし、病院のがん診療の負荷、すなわち、年間にがん患者が何人診療されたか、などを評価するための作表には含めてもよい。

(1)から(6)までの範疇が全症例を「網羅」しているわけではなく、がん患者の中には、全く登録に含められないものもあるという点に留意すべきである。診断や治療計画を確認するため、相談にだけ来院した患者や、例えば、休暇や元の病院での装置故障などのために、他で始めた一連の治療を中断しないために、一時的に医療を受ける患者、がんの既往歴があるが、現在は疾病の兆候がない患者、などがそれである。

病院によっては、性状不詳の新生物、良性の病変、特に良性の脳腫瘍、前がん病変をも登録したいと願うところもある。院内登録には、国際疾病分類（WHO, 1976b）（ICD-0）の性状コードの /2または /3で表されるすべての病変を含めることが推奨されている。この規則の例外は、皮膚の基底細胞がんと扁平上皮がん、子宮頸部の上皮内がんの登録である。多くの病院では、これらの症例の登録は、（症例数が多いため）非常に時間と費用がかさむことが分かったため、登録処理から除外する傾向にある。アメリカ外科学会の勧告（ACSCC, 1986）では、皮膚の限局性の基底細胞がんと扁平上皮がんを除外することとするが、診断時点で領域浸潤があったものは含めるべきとしている。さらに、子宮頸部の上皮内がんは、患者索引ファイルに入力されるが、データベースには、完全には抜粋する必要がないとされている。

伝統的に、殆どどの院内登録では、抜粋の完了のあと、症例要約票を、一定の年次別部位順にファイルするという、手作業による操作を行ってきた。しかし、次第に、院内登録は電算化されてきており、データが単に要約されるだけでなく、データがコード化され、キーデータも入力されるようになってきている。手作業とコンピュータによる登録の殆どは第8章で述べられており、以下の記述は院内登録に、より重要な点に絞っている。

症例の把握

病院内部には、いたるところでがんの診断がなされ、記録される場所がたくさんある。それ故、これら情報源の逐一を同定し、適切な記録を入手しうよう、手配することが必要となる。多くの国では、さまざまな記録システムの所有権について、また誰がそれらの記録を閲覧でき、あるいは出来ないのかについて疑問があり、事態は複雑である。しかしながら、施設内の全ての記録システムを同定し、それから必要な情報を入手しうよう手配することは、明らかに、それぞれの病院の腫瘍登録士の責務である。以下の事項については、熟慮が必要である。

- 血液、血球検査の記録と細胞診の記録とが別の部門で保管されているのか、もしくは、そうした全記録が病理記録の中に収納されているのか？—
- 剖検記録はどこで管理されているのか？—
- 入院患者記録の調査では決して認められることのない症例を検索するために、外来患者記録を使用しているのか？—

大抵の病院において、症例同定のための2つの主要情報源は、病理記録と診療録部門疾病索引である。多くの場合、疾病索引はコード化され、コンピュータ化されているので、がんのコードを付された症例をリストに出力し、これを活用することができる。顕微鏡的に確認されなかった症例については、「probable」、'possible'、'consistent with'などの様な、

非特異的な用語が最終診断の一部になっている時に、臨床的に診断されたものの内、如何なる症例を含めるのかを決定しておかなければならない。次に示すのは、登録に含めるか、除外するかを決める場合に、慣例的に使われる用語のリストである。

- あいまいな表現の「'probable'、'suspect'、'suspicious'、'compatible with'、または、'consistent with'」は、腫瘍による障害と解釈される。
- あいまいな表現の「'questionable'、'possible'、'suggests'、'equivocal'、'approaching' または、'very close to'」は、腫瘍による障害と解釈されない。

登録の過程

院内登録の実際の登録過程は、第8章で述べた原則と少し異なっている。もし、カードの形でファイルを維持するならば、それらは、登録番号簿、患者索引ファイル、腫瘍記録ファイルから構成される。登録がコンピュータ化されると、もちろん、登録番号、患者名、腫瘍部位などによって、データベースへのアクセスができる。

一旦登録されるべき腫瘍が同定されると、患者と腫瘍に関する情報が、診療録から、予め決められた様式を用いて、あるいは書式抜粋の中間段階を経ることなく直接コンピュータへと、抜粋される。第8章で述べた様な、コード化と媒体変換についての考慮は、ここに関連している。多くの院内登録は、比較的少数の症例（地域登録と比較して）しか情報を記録しないので、コードと同様、記述情報も、コンピュータに入力し、コード化された情報の記述化を、幾分かしておくことが推奨される。多くのデスクトップコンピュータは、大きな記述情報を維持するのに適した記憶容量を持っていないので、記述情報は、一度入力されると、記録票/要約票として印字され、手作業によるファイルの形で維持され、コンピュータ化された記述情報は消去される。原発部位、組織型、病変範囲のような項目の文章記述は、精度管理の目的、及び将来の研究用のより詳細な情報保持のために、欠くことが出来ない。後者の例をあげると、患者はコンピュータ化されたデータベースの中で、ICD-0の部位コード T-173.6（腕及び肩の皮膚）として登録されるが、記述部のバックアップは、病変が、手、手掌、手首、前腕、肘のどこにあるのかを区別することができる。

診療録の記述による要約票を維持するもう一つの理由は、がん患者記録を用いる特別な研究が実施される時に、病院診療録がしばしば手に入らないからである。病院診療録は、もし患者が何かの理由で再入院した場合には他でも使用されるかも知れないし、患者が死亡した場合には取り出しが困難な状態で保管されているか、もしくは、患者が死亡して一定の期間がたてば、処分されている可能性もある。それ故、院内登録内で維持される関連情報の抜粋は、欠くことの出来ないものである。

抜粋に含まれる項目は、病院と病院がん委員会で決定される。しかし、少なくとも、理学

的検査と病歴、診断のための検査、臨床病理検査、病理報告及び手術報告に関する必要事項を記入する為のスペースを設けるべきである。治療の詳細は、病院によって特定されたレベルで記録されるべきである。しかし、最小限、患者が、手術、放射線治療、化学又はホルモン療法、免疫療法、またはその他のあらゆる認定された治療法を、受けたかどうかを記録するべきである。化学療法に関して、しばしば薬剤が症状緩和の為にだけに使用されるので、特定の薬剤は精度管理の目的で、記録されるべきである（例えば、プレドニゾロンは根治の目的よりは、抗炎症剤として使用される）。

抜粋が完成し、確認されると、将来の利用のためにファイルされる。もちろん、コンピュータ化された登録は、ファイルする前に入力する。上述した様に、登録は、全ての記録を電子工学的に維持することを選んでよいが、コンピュータの蓄積容量の限界が、そのような計画を不可能にすることもある。アメリカ外科学会は、抜粋を、部位別、診断年別、患者氏名のアルファベット順に、ファイルすることを勧めている。これにより、部位別の統計的総括をする場合に、抜粋が利用され易くなる。患者数の多い病院では、死亡した患者の要約票は、しばしば、分離してファイルされたり、マイクロフィルムやマイクロフィッシュの形で保管される。

多くの大病院では、患者の退院後しばらくの間、診療録が抜粋に利用できないことがある。すべての抜粋は、可能ならば常に患者の退院から6カ月以内に完成することが勧められている。また、抜粋は、必要な情報がまだ患者記録の中に含まれていないことがあるので、退院後余りに早く行うべきではないと勧告している。しばしば、臨床病理報告、手術報告、及び病理報告は、退院後すぐには利用できないことがある。

正確さと一貫性が最も重要な点であるので、再抜粋を含む、そして、もしコンピュータ化されているのなら、記録の一部抽出サンプルを再コードする精度管理プログラムを備えているべきである。病院登録士の訓練の継続と再訓練は、精度管理プログラムの欠くことのできない部分である。

データ項目

先に述べたように、院内登録は、病院管理と患者機能とに対応してゆかねばならないので、特定人口集団におけるがんの罹患を計測することを主目的とすることが多い地域登録にとって、興味のない項目をも含んでいる。地域登録に関心のある項目は、第6章で述べたので、ここでは繰り返さない。

院内登録にとって重要な次にあげる項目は、先には述べられておらず、これらの重要性は地域登録にとっては低い。

—配偶者、友人、保護者の氏名

- 電話番号
- 病院の診療科
- カルテ番号
- 入院日
- 退院日
- 紹介元
- 紹介先
- 主治医
- 他の担当医
- 症例区分（定義は先に示した）
- 診断方法
- 病変範囲（TNM分類；腫瘍の大きさ；検査したリンパ節の数；陽性リンパ節の数；サマリー・ステージ）
- 初回治療の日
- 初回治療：
 - 外科手術、術式と範囲を含む。
 - 放射線、照射結果
 - 化学療法
 - ホルモン療法
 - 免疫療法
 - 他の治療
- 残存腫瘍、遠隔部位
- 初回再発の日、様式、部位
- 後療法の日、型
- 退院時状況
- 初回治療の（1）前（2）後、及び、記念日での患者の状態（quality of life）
- 連絡先住所、氏名
- 経過観察医

病院管理の目的で、院内登録は、施設の利用状況を調べることに関心を持つかもしれない。それ故、在院日数（入院日と退院日との対比）の計測以外に、登録は、装置の利用、更新、アップグレード、あるいは廃棄を正当化するために、病院当局を支援する目的で、CTスキャナー、生物学的マーカーの検査、表現型タイピング、電子顕微鏡、エストロゲン感受性測定、などの利用状況を調べることもある。また、紹介（紹介元、紹介先）のパターンを調べることによって、ある病院の診療圏を、より明確にすることが出来る。このことは病院管理

者にとって、他の病院との協力関係を設立する上で、また、多施設での訓練と継続教育プログラムとを立案する上で、さらに、装置と資源を共有する上で、非常に役に立つ。

患者のフォローアップ

院内登録の主要な関心事は、患者の健康の保持と医療とにあるので、患者に便宜を供給する目的で、以下に示す追加関連項目を記録するべきである。すなわち、退院後、どの医師が患者に対して責任をもつのか（外科医、腫瘍専門医、または開業医）；患者が他の病院へ紹介されたかどうか、そして、もしそうなら、どの病院へ；退院時の患者の状態（QOL）、及び、時間の経過とともに患者の状態がどう変化したか、そして、いつがんが再発したか。

患者をフォローすることは院内登録士の責任となるので、2、3の接触のポイントを確立しておくべきである。主要な接触のポイントは、患者の医療に責任を持っている医師を通じて行うことである。しかし、どの医師が主に責任を持っているかは、必ずしも明らかでなく、加えて、患者が医師との接触を失うかもしれないし、医師が転居、退職、死亡する可能性もある。そこで、院内登録士は直接に（現住所、電話）、もしくは配偶者、保護者（小児の場合）、親戚、友人を通して、患者に接触する方法を知っておくように努めるべきである。特別の腫瘍型の場合は、患者は決まった頻度で（6カ月ごと、1年ごと、など）接触が保たれるべきである。しかし、疾病の性質から、また、そのための時間と出費が大きくなるために、子宮頸部の上皮内がん患者をフォローすることは、勧告されていない。フォロー時に、患者の疾病と機能状態、その後治療を受けたか、もしそうなら、どこで、また、何時、最後に医師の診察を受けたか、そしてどの医師か、を記録するようにする。これらの項目は、その後、患者の疾病の進行をモニターし、次のフォロー時に患者と連絡する際に役立つ。

患者のフォローに登録室がどれだけ成功したかを示すのに、さまざまな基準が使われている。最も一般的な指標は、全登録例についての成功率の計算である。すなわち、例えば、1987年の成功率を計算するためには、1987年及びそれ以降に最終接触をしたすべての生存患者と、死亡の判明している全患者とを計上し、その合計を全登録例で除す。この率は、その登録室の追跡成功率を表す。この計算方法は、これまでどれだけ良好なフォローアップが行われてきたか、また、その様なフォローアップに基づいた生存率計算が、いかに正確であるかを示す良い指標となるが、症例数が多く、かつ、長い歴史を持つ登録室にとっては、最近年のフォローアップが、それ以前よりずっと低率となる可能性があるために、この指標は誤解を招くことがある。

この点は、ある例によって大変よく示される。すなわち、ある病院は、1978-87年の10年間に毎年1,000症例を登録し、その結果、全部で10,000人が登録された。1987年末の時点で、7,000症例は既に死亡していることが判明し、2,200名は1987年のある時点で接触があった。

(1987年の登録例は、定義により1987年のある時点で全例、接触があったということになる。) この登録室でのフォローアップ成功率は、7,000(死亡例)と2,200(1987年接触例)を加えた数値を、登録数の10,000で除して得た92%ということになる。しかし、もしも7,000の死亡例の内6,400例は1987年1月1日以前に死亡しており、それ故、1978年1月1日から1986年12月31日の間に登録された9,000人の内2,600人は、1987年1月1日時点でなお生存していると考えられ、従って、その病院のフォローアップ必要数は、1987年中に接触する必要のある2,600の既診断例ということになる。さらに、1987年に接触された2,200例の内、もしも1,600例は1987年以前の診断例で、600例が1987年の診断例であったとし、そして、もし、1987年に死亡した600例の内200例が1987年以前に診断された例で、400例が1987年に診断された例であったとすると、その場合の、1987年に登録室がフォローしなければならない2,600症例についての成功率は、実際は、1,600(1987年に接触できた生存例)と200(死亡例)の合計を、フォローしなければならない2,600で除して得た69.2%ということになる。この成功率は、直前の12ヶ月の間に達成し得たその登録室のフォローアップ機能が、どうであったかを評価するものであり、この率は、病院のフォローアップがどれほどうまくいっているか、最近の患者がどのようにフォローされているかを評価する、より正確な指標である。フォロー成功率の計算は、この第2の方法で実施されるよう推奨する。そして、院内登録の目標は、少なくとも90%のフォロー成功率を獲得するように勧告される。

データの報告

すべての院内がん登録は、登録データを、病院理事会と病院がん委員会とに、毎年報告すべきである。報告書は作成されるべきであり、アメリカ外科学会は少なくとも下記の事項を含めるべきであると勧告している。

- この病院の対がん医療計画の目標、成果、及び活動についての記述要約；
- 登録室の活動に関する報告
- その年の登録データの統計サマリー。これには、原発部位の分布、主要部位に焦点を当てた表やグラフ、そして、フォローアップ活動に関するデータを含め、データと病院内におけるがん患者の管理とを結びつける短い記述を添えるべきである。
- 1つもしくはいくつかの主要部位についての、詳細な統計的分析。これには、生命表方式によって計算した生存率データ、適切なグラフ、表、記述で示される他の記述統計、そして、がん委員会の医師メンバーによるデータの総括的論評を、それぞれ含むべきである。

年次報告に加えて、登録データは、全ての腫瘍委員会とカンファレンスとで利用されるべきである。さらに、病院がん委員会は、全病院スタッフが、適切にこのデータを利用するよ

う勧奨すべきである。また、院内登録士も、独自に研究を始めるように、そして、年次推移における通常でない変化を指摘し、これらの変化が何を意味しているのか、がん委員会に疑問を提起するよう、勧奨されている。

病院レベルでのデータの活用が、そうした活動の経費を正当化する唯一のものである。要約すると、最高の精度のデータが利用できるように、登録の手法が適切に、正確に記録され、手法が遵守され、時宜を得た方法で症例の同定と登録が行われ、さらに登録室の利用のために診療録からの情報が正確かつ完全に抜粋されることを保証することが、がん委員会と共同して働いている登録士の責任である。

(翻訳：北川貴子、津熊秀明、陶山昭彦、服部昌和)

Chapter 14 Cancer registration in developing countries

D. M. Parkin and L. D. Sanghri

第14章 発展途上国におけるがん登録

〔 省 略 〕

第15章 がん登録：法的側面と情報保護

C. S. Muir and E. Demaret

序言

誰もが医療を受ける資格を持っている、ということは明らかである。健康・医療 (Health care) は単に疾病の診断と治療のみでなく、疾病の予防、対策、及びリスクの測定 (がん登録による) をも含む。世間が情報保護の問題にめざめ、医学ファイルとその他のファイルとをつなぐようなことを含め、個人のプライバシーに関し警戒心を持つようになり、医学及びその他の個人情報を保護することが、ますます重要となってきた。従って、個人のプライバシーの権利を守る法律が、多くの国で作られつつある。この方向で幾つかの国の政府がとった方策は、しばしば、がん登録データの最適利用とは調和しない。むしろそれを妨害するような情報保護の方策を負わせる法律及びデータ保護条例をつくる結果となってきた。

データの保護は、データを保存し、保護すれば大きな疫学的価値がある場合に、データを保護せず、むしろ破壊を促進していることがある。偶然、或は審議の上、破壊を免がれた多くのデータファイルは、その後の研究で非常に重要なものであることを証明したが、他方で依然、そのようなファイルが悪用される、と宣伝されるままとされている。がん登録のデータは、決して価値を失わず、40-50年間も前に収集されたデータが、なお高い頻度で利用されている。

がん登録の法的根拠

がん登録は、患者についての届出に基づいているが、それは篤志によって行われる場合と法的な制度による場合とがありうる。

法定による届出は、保健省又は州の衛生部が定めた法令により、又は行政命令 (指導) に基盤をおいている。地域によっては、届出は、施設の種類により、篤志と法定との両方の届出であることがある。例えば同一の登録地域でも、病理医は篤志により報告するが、人口動態課は報告を義務づけられている。又別の地域では、病理医は法律でがんの報告を要求されている一方で、治療医は卒先し、自発的に届出を行っている。がんが法的に報告されるべき疾患でない地域では、がん登録は、実際には、篤志に基づいて運営されている。いくつかの国では、同定可能な形で情報を登録室へ篤志報告するには、明示された患者の同意を必要とする、ことを意味することになる。

篤志による届出が、必然的に不完全な届出を意味するものでないことを強調しなければならない。多数の国で報告は今なお篤志によっているが、幾つかのこのような地域のがん登録

は、義務的に報告をうける地域と同等、又はより高い質で行われている。最近10年間の技術分野での発展によって可能となった、社会の個々のメンバーに関する情報の幅広い収集と利用は、プライバシーと情報保護の問題への国民の意識を高めた。この状況下で、がん登録についての法的側面はますます重要となってきた。

しかし行政的及び法的規定は、登録を助ける面と妨害する面と両面を持つ。法律によりがんを届出すべき疾病にすることは、届出を増やし、関係者、関係機関が届出を行う位置にあることをはっきりさせるかも知れない。しかし他方、プライバシー保護法は、同定項目の登録を不可能にしたり、がん患者が登録を拒否するようにさせるかも知れない。例えば、がん登録に届出を行った場合に起りうるべき結果を医師が懸念したことによって、ハンブルグがん登録では、1979年以前には年間平均1万件あった届出が、1980-81年には2件しかなかった。これはがん登録室と、大規模な届出機関でもある厚生省との間での、情報の送付に関する規則が改訂されたことから生じた。即ちこのような協力態勢のための法的根拠が存在していなかったためである。新法が1985年に発効し、医師ががん登録室へ症例を届出することができるようになったものの、同意をえてから届出をすることになった。

がん登録室とそのデータの利用者は、がん登録室が高度な質のデータを収集し続けること、また、そのデータをもっている人の匿名性を保ちつつ、可能な限り効率よく利用されること、が保証されることに対し、強い関心を持っている。

情報保護の目的

第3章に記述したように、どのがん登録事業の目的も、集積した、または個々のデータを、医学研究と統計目的のために利用することである。価値が高いものとするためには、記録されたデータは可能なかぎり正確で、質的精度が高く、量的にも完全なものでなければならない。登録の過程と、データを最大限に利用する面との両方で、個人の同定が不可欠である。正確性と完全性とは、一重に、公衆と治療医が、要請されているデータが登録の目的と目標とのために必須であること、及びデータは保護されていること、を信じることによって達成できる。がん登録室でのデータの保護とは、単にデータが許可されていない者は利用できないように十分に安全に守られているのみでなく、それらが収集される目的以外の目的には使用されないことを意味している。

がん登録事業における情報保護の目的は、以下の事項を保証することにある。

- a) 登録室に報告された患者個人の名前が外部に知られないように保護し、もし必要なら報告者の名前をも同様に保護する。
- b) がん登録データは可能なかぎり最高の質のものとする。
- c) がん患者の利益、がん対策、及び医学研究のためにがん登録データを、最大限に利用する。

情報の保護

がん登録室に報告されたデータの匿名性を保護するために、がん登録室が保護規約を作成するようすすめる。情報保護はがん患者データに適用されるのみでなく、その他の者、例えば、登録室と関係をもっている業者、或はその他の集団に対しても適用され、その情報が死亡者のものであれ生存者のものであれ、区別されない。多くの登録室は、届出医師名は病院の届出科名が出ないように、保護している。

登録事業の長は情報の保護に対し責任を持つ。この責任は適切な法律又は行政命令によって定められるべきである。しかし情報保護は登録事業の中の全スタッフにかかわるものである。そこで、雇用時に、スタッフに、がん登録におけるデータのどのような情報も口外しないという特別の宣誓書にサインしてもらうようにすすめる。この誓約を破れば、懲戒処分となることを明記すべきである。この情報保護の宣誓は、退職後にも継続するものであることを強調することが重要である。情報保護の必要性に関する注意が、登録室の中で掲示されるのもよい。また、がん登録では、正式に適切な期間において、情報保護処置の見直しを行うことをすすめる。

がん登録における情報保護の実際

情報は定まった道を通って登録室に到着し（第5章参照）、通常それは作業規則（第6、7章参照）に従って処置され、報告書又はその他の出力が作成される（第10章参照）。何種類もの出力が、内部の使用目的のためにのみ、つくられる。登録の作業過程のフローチャートを作っておくこと、そのどこで情報保護を保証するための方策を適用すべき必要があるか、を決めておくことは有用である。

特別な配慮を必要とする項目や、実施されてよい何種類かの処置を以下に述べる。

届出票の収集

がん患者の届出は、治療医、病院病歴室、退院係、病理、細胞診、血液学及び生化学の実験室、放射線医、検視官及び人口動態統計局（死亡診断書）などの多くの情報源から行われている。これらの報告は、普通、患者名やその他の同定情報を含むが、これらの内容は、情報源と登録室以外の人々の目に触れることはない。もし登録室のスタッフが情報源から情報を集める（出張採録など）場合は、彼らはがんに関する情報に関してのみならず、彼らがその情報源の施設で見たり聞いたたりしたすべての機密性を持つ事柄について、その機密を守る責任を負う。その施設で登録スタッフが集めたデータを輸送する際には、鍵がかかるアタッシュケースを用意する、などの考慮がなされるべきである。

源施設から登録室へ、或は登録室から、の情報の輸送

情報は、郵送、テープ、ディスク、コンピュータ・ターミナル又は電話によって転送されることがある。それぞれでの注意を、次に記す。

① 郵送

考慮されてよい安全処置の中に、次のようなことがある。

- a) 書留郵便 b) 患者の名前のリストと他の情報を別々に送る c) 特別な封筒を使わない
- d) ダブル封筒を使用し、外には普通の宛名を書き、中には〇〇親展と記しておく。

② 磁気メディア

情報が磁気テープ、ディスク又は他の類似のメディアで送られてくる時には、

- (a) 行先不明で途中で紛失しない (b) 第三者が簡単に読めない
- (c) 許可なく登録の施設内に置き去りにしない

などを保証するよう予防策が取られるべきである。

予防策では、氏名の記号(暗号)化がある。これは種々なレベルの複雑な形で行われうる。また①名前と住所用と②腫瘍関連データ用とに別々のテープ又はディスクを、用意し、リンクのための番号を組み入れておき、氏名テープを最大限保護する。発送した全磁気テープ、ディスク、又は他のデータメディアの記録を、登録室が受け取るまで残すこと、をすすめたい。

③ 電算機

情報は電算機を介して送られてくるかも知れない。また登録室は、この情報を、外部にある電算機の記憶装置へ送るかも知れない。この時とるべき予防策の中に、利用者の同定とパスワードの利用(パスワードは入力されてもVDUの画面に出さない)があり、また、許可されて行われた入力のための電算機利用時間を記録すること、利用者が所定の日誌に書かれた情報をチェックすること、などがある。パスワードは時々変更すべきである。転送時にも氏名は暗号化するよう考慮されねばならない。情報システムの発展に伴い、量的に増加するデータは、がん登録室へ公共電話又は専用電話回線で送られることになってゆくだろう。

④ 電話

時々、源施設から、又は登録室から情報をえるために、特に不明の項目を完成する場合に、電話が用いられる。便利ではあるが、電話は機密保護の侵害を発生させやすいことを認識していなければならない。電話を受けた者が認定された受話者でなく、さらに自分自身を同定する証拠をも述べない時には、機密情報をあたえるべきでない。

データの使用と保管

(1) がん登録室

許可されていない者の登録室への立ち入りは、禁止されねばならない。登録室へ入室できる者の名前のリストを作成するように勧める。必要な制御、施錠、及び警報ベルシステムが、装備されるべきである。

(2) 電算機

がん登録室の情報の大部分は電算機に保管されているが、電算機の手ほどきを受けていない者は、これに容易に触れないものである。電算機内のデータへの接触は、普通はパスワードの使用によって保護されている。さらに安全な処置は、氏名ファイルをその他の情報から分離しておくことである。後者には、氏名ファイルにもどるための特別のキー又はパスワードをつけておく。

登録室長は、登録スタッフのリストをつくり、その中の各自が接触できるデータの種類やレベルを定め、更新してゆくようにすべきである。その日の作業の最後には、外部保管用のメディアは、鍵をかけた耐火金庫に保管すべきである。

(3) カードファイル

大抵の登録室は、紙に書かれたデータを相当量持っているが、これは容易に人に読まれうる。この材料には、例えば届出票、内容を抽出すべき症例記録、病理所見のコピー、死亡票のコピー等々が含まれる。このような材料で氏名と他の情報とを分けて保管するのは実際上、不可能である。従って、これらの情報は可能な限り安全に保つことに配慮せねばならない。考慮されてよい処置の中には次のものがある。

- (a) 登録室の施設内へ入りうる人を定めておく。
- (b) 資料が保管されている部屋に入りうる職員をきめておく。
- (c) 1日の仕事を終了した時に、資料を入れるべき施錠可能のキャビネットを用意する。
- (d) 許可されていないスタッフ（例えば、清掃員）が記録類コピーや廃棄された用紙を含む一をほじくりかえすことができないようにしておくこと。

(4) 死亡者のファイルの処置

多くの登録室では、カードファイルは、登録患者が死亡したことがわかってから約2年間程度、保管する。これらのファイルは、その後マイクロフィルム化し、フィルムは無期限に保管する。原票は廃棄する。普通、裁断又は焼却により、廃棄する。

(5) 登録活動の中止

どの登録室も、活動をやめる場合にとるべき行動に関し、方針を持っているべきである。中止時には、登録室内の全記録をマイクロフィルム化し、適切な組織において、最短35年間

保管することをすすめる。その間、がん登録室が活動していた時と同じ情報保護の規則を遵守する。

データ利用と貸出し

もし登録室が要請に応じることになっているのなら、そのデータは利用のために貸出しされねばならない。データがそのために貸出しされる目的の中には、しかしながら、情報保護の問題が持ちあがる。

データは、がん登録室では文書による要請時のみ、用意されるべきである。要請は、次の事項を含めたものであるべきである。

- a) データを必要とする正確な目的、
- b) 必要な項目、
- c) 情報を保管する責任者名、
- d) データを必要とする期間。

データが要求された時には、次の事柄を確認しなければならない。

- a) データを受け取る者は、登録の職員が遵守した情報保護に関する原則によって拘束されねばならない。
- b) 受領者は、データ利用時の制限に従うこと。特に計画時に同意をえた以外の目的のためにデータを使用しないこと、及びデータを第三者に伝達しないこと。
- c) 患者およびその家族との接触は、主治医の許可をえた後にのみ許されること。
- d) 指定目的に使用する必要がなくなった時には、データは返却又は破棄されること。

診療目的による利用

診療目的のために医師に対し、データを提供することがあるかも知れない。診療はチームの協力の下に行われることがますます多くなっているが、このことは、情報の保護は、チーム全員により保護されることを意味するものであり、データ公開のための規則では、このことを考慮に入れておかねばならない。

(1) 統計及び研究目的の利用

統計及び研究目的のためのデータの利用および貸出しは、個々の人の利益となる知識を進歩させること、保健サービスを改善すること、又、保健および保健行政と対策立案を援助することを目的とする。

(2) 集合データ

がん登録室が行いうる最も重要な貢献の一つは、がん罹患の最近のデータを、年齢、性、出生地、職業別等に作成することである。組織型の分布差、都市と郡部の差、時間的推移等が検討される。このような作表の時に、機密保持上、問題のおこりうることは、殆んどな

い。表中で個人を同定することは、一つの枠の中の患者数が非常に小さい場合には、潜在的可能性としてあるかも知れないが、このことが実際になされる危険は非常に少ない。しかし表には、このような危険を最小にするよう計画されるべきであり、登録のルチンの報告書に出す詳細データのレベルを考慮しておくべきである（例えば各枠中の数、届出施設の同定、届出施設別生存状況、地域別率など）。

（３）個々のデータ

ケースコントロール研究やコホート研究は、がんの原因を同定することを助ける。登録室は、このような研究に重要な貢献をしている。そして、ケースコントロール研究に含めるために、登録室は、例えば、あるがんを持つ人々の名前を提供することを求められるかも知れない。もし登録参加医が各患者から同意を与えられていないならば、患者名は洩らされるべきではない。逆に、予め登録参加医又は参加病院の科の許可を受けることなく、患者やその他のどの人々にも接触しない、という条件をつけて、善意の研究者に患者氏名を示してもよい。研究者の大多数のためには、匿名又はグループデータの報告で十分である。出版された疫学的研究では、研究で扱われた個々の患者の氏名を一度も洩らすことはなかったことが強調されるべきである。

（４）国際的資料提供

データは、通常、個人が同定されうる形で、外国に送られることはない。もとの国で、データの確認のため症例にもどる時のために、対象一人一人に連続番号又はその他の名称を付与する。個々の患者は、この番号などによって、必要な場合には、もとの登録室スタッフにより、もとの登録室で捜し出されうる。研究状況から必要となった時には、国の法律は、個人データが国境をこえることを許可する。例えば移民の研究など。このような研究のデータでも、データを用意した国の機密性の規則に従って扱われるべきである。

（５）管理目的

個々の情報は、生命保険、疾病保険、年金制度その他の同様の管理的目的のために提供されてはならず、このような目的のために健康診査をする医師にも提供されてはならない。

（６）公的組織、新聞及び一般大衆に対する定期報告中のデータの配布

データは、統計である間は、年報等の報告を通じて配布されるが、その結果、データは匿名のまま、個人を同定できない形で提出されるべきである。

がん登録室は、しばしば新聞によって様々なトピックスに関する情報を求められる。そこで、登録室の中で指名された人が、このような調査に対応することを勧める。

がん登録室は、しばしばシステムのデモンストレーションを頼まれる。デモンストレーション実行時には架空のデータを使い、見学者が気付くよう、“デモ用”とラベルをはっておく。このような機会のために特別のパスワードを用いることは有用であるかも知れない。

レコードリンケイジ

がん患者の記録を個人別にリンクすることは、がんを正確に計測し（第8章参照）、がんの原因を最小の支出で確かめ、対がん施策の効果を評価する、などのためには、必須である。リンケイジでは、各レコードが同定に関する情報を持っていることを必要とする。外部ファイルとのリンケイジの目的は、（a）生存率計測のためのフォローアップ、（b）治療結果をみるためのフォローアップ、及び（c）疫学研究、を実行するためなどであろう。照合が登録室内、或は室外のいずれで行われるにしても、データの機密性は守らねばならない。同様のことが、照合において登録室が同一人かどうか判定不明とした患者が発見され、届出施設から、より多くの情報をもraitたい時にも適用される。

もし照合が登録室の外、例えば人口動態課で行われねばならない時など、または、第3者に属するコンピュータを使用する時などには、登録室は、登録情報が保護されること、及び登録データを受け取った組織が情報保護規則を同等に遵守することを、確かめねばならない。同様に、もし外部のレコードとの照合を登録室内で行う時には、同一の情報保護の規則を適用すべきである。

登録ファイルと死亡票との照合によって、死亡票のみ（DCO）の症例が同定され、登録室がその症例の情報をさらに探す場合、死亡票を作成した医師に要請する。もし国が、規定している場合は、この要請は人口動態課を通じて行うことになる。

もし他ファイルと照合時に、登録がこれまで未登録の患者の存在を疑った時には、相手の組織自身が情報保護の規則によって閉ざされているならば、登録はそのデータファイルに責任のある組織に申し入れをし、情報をさらに得るようにする。

電算機への無許可での侵入

情報を盗むため、或はもっとしばしば、単に侵入することが可能であることを示すために、電算機へ侵入しつづけた人の例がある。しかし、著者らは、がん登録室に、こうした例が起ったことを聞いたことがない。がん登録室は、そのシステムを完璧に守ることができるとは思われないから、うちたてべき安全性のレベルは、侵入を可能とする普通の方法の裏をかくようなものであるべきである。

がん登録室専用の独立したコンピュータが機密保持力を増す。

要約

とられるべき安全方策をあげると、次のごとくである。

a) 登録室への入室者は、限られた資格のある者のみとする。

- b) 電算室への入室者は、限られた資格のある者のみとする。
- c) 電算機の取扱者は、情報に近づきうるパスワードを与えられた、限られた資格のある者のみとする。
- d) VDUの画面に出なくしたパスワードとユーザーキーを使用する。
- e) 許可をうけた各人の電算機利用時間を記録する。
- f) 氏名ファイルは他の情報ファイルから分離し、氏名は暗号化又は乱列記させる。
- g) 情報を輸送する場合、スタッフに、鍵のかかるアタッシュケースを使用させる。
- h) 登録室スタッフを同定する方法を用意する。
- i) データを収集し、転送し（郵送、テープ、ディスク、電算機或は電話で）、保管し、データを外部に出し、あるいは照合する時に、データを保護するために、特別の注意を払う。
- j) 個人の同定が可能な全出力に対し、コントロールする方法を講じる。
- k) 登録ファイルと外部ファイルとの照合の権利の制限。
全体を通し、登録室長は、高レベルの安全を守り、情報を保護する必要性を、スタッフに吹き込まなければならない。

結論

幾つかの国では、公衆は情報保護の問題に目覚めてきている。特に電算機と電算機中のデータ記憶装置の幅広い利用を通して、関心は「電算機中の氏名」についての懸念につながっている。

特別な知識と電算機を使う権利とをもった人々のみが電算機を動かしようというように電算機を閉鎖すれば、氏名はそこでは、カード上にあるよりはもっと安全である。電算機による暗号化の技術は、今では実用的には、破ることが出来ない程になっている。がん患者の氏名と非医学ファイルとの照合は、本格的な社会的関心事となっている。北欧の国々は研究ベースでの照合を承認するデータ保護の理事会を持っている。英国では英国医師会の倫理委員会の承認が必要である。研究者が研究目的を説明した後で与えられることになるこのような公開の許可制は、乱用を制し、研究の質を高めるものと思われる。

がん登録は長い間、がん患者と社会医療に相当な貢献をしつつあるとみなされて来た。そして、そうすることの中で、自身の情報保護の規約を保持してきた。多くの登録は、しかしながら、書かれた規約を持っていない。がん登録がここに概略をのべた一般原則を登録地域の状況に適合させた形で、自身の規則や規定の草稿を作成することを勧告する。

がんについて知っている人々、すなわち患者と届出医のみ、の間で秘密を絶対的に守ることとは、実際には、各がん患者ががんを持つ他の人々の経験からの恩恵を受けること、及び、この疾病についての知識を集積するのに貢献すること、を妨げることになる。このような秘

密蔵守は、産業等の危険が隠されたままになったり、或は故意に隠蔽されるままとなることを容易にし、また、治療、スクリーニング及び予防等の対策における知識の蓄積から受ける恩恵を共同体が評価することを妨げる。著者らは情報の保護を履行しなかったがん登録を知らない。倫理的ながん登録であっても反対し続ける人々は、重い責任を負うことになる。

(翻訳：花井 彩)

付録 1 . 国際連合標準国別コード

国際使用のために構成された数字コードを、以下に示す。各コードは、国・地域を表わす。基本数字コードは3桁の数字より成り、国・地域を特異的に識別する。この数字は以下に示すように、英語表記国名のアルファベット順に並べられている。基本数字コードに間隔があるのは、リストの将来の発展、拡大を予想したからである。本リストに挙げられた国名・地域は Country Nomenclature for Statistical Use, 第7版 (国際連合統計局文書) に基づいている。

分類の基本単位

コードを付した範疇は、国・地域の地理的領域を表わす。また本コードは、場合によっては国籍を示す目的で用いることもできる。例えば、250のフランスはフランス人、504のモロッコはモロッコ人を示すこともできる。

本分類は国・地域を完全にカバーしているが、本分類表の利用者が定義できる特別の範疇が存在することをも考えて作成されている。この要求を満たすため、900～999のコードには割り当てがなく、利用者自身の目的のために保留されている。本分類表を利用した資料のあらゆる伝達において、9で始まるコードを単純にチェックするだけで、すべての非標準コードを除くことができる。

(翻訳：藤田委由、津熊秀明)

000 Total	052 Barbados
004 Afghanistan	056 Belgium
008 Albania	058 Belize
012 Algeria	204 Benin
016 American Samoa	060 Bermuda
020 Andorra	064 Bhutan
024 Angola	068 Bolivia
028 Antigua	072 Botswana
032 Argentina	076 Brazil
533 Aruba	080 British Antarctic Territory
036 Australia	086 British Indian Ocean Territory
040 Austria	092 British Virgin Islands
044 Bahamas	096 Brunei
048 Bahrain	100 Bulgaria
050 Bangladesh	854 Burkina Faso

104 Burma	234 Faeroe Islands
108 Burundi	238 Falkland Islands (Malvinas)
112 Byelorussian Soviet Socialist Republic	242 Fiji
830 Cameroon, Republic of	246 Finland
124 Canada	250 France
128 Canton and Enderbury Islands	254 French Guiana
132 Cape Verde	258 French Polynesia
136 Cayman Islands	260 French Southern and Antarctic Territories
140 Central African Republic	266 Gabon
148 Chad	270 Gambia
152 Chile	274 Gaza Strip (Palestine)
156 China	278 German Democratic Republic
162 Christmas Island	280 Germany, Federal Republic of
166 Cocos (Keeling) Islands	282 German Democratic Republic Berlin
070 Colombia	284 Germany, West Berlin
174 Comoros	288 Ghana
178 Congo	292 Gibraltar
184 Cook Islands	296 Gilbert and Ellice Islands
188 Costa Rica	300 Greece
384 Cote d'Ivoire	304 Greenland
192 Cuba	308 Grenada
196 Cyprus	312 Guadeloupe
200 Czechoslovakia	316 Guam
116 Democratic Kampuchea	320 Guatemala
720 Democratic Yemen	324 Guinea
208 Denmark	326 Guinea-Bissau
262 Djibouti	328 Guyana
212 Dominica	332 Haiti
214 Dominican Republic	336 Holy See
218 Ecuador	340 Honduras
818 Egypt	344 Hong Kong
222 El Salvador	348 Hungary
226 Equatorial Guinea	352 Iceland
230 Ethiopia	356 India
	360 Indonesia

364 Islamic Republic of Iran	496 Mongolia
368 Iraq	500 Montserrat
372 Ireland	504 Morocco
376 Israel	508 Mozambique
380 Italy	516 Namibia
388 Jamaica	520 Nauru
392 Japan	524 Nepal
396 Johnston Island	528 Netherlands
400 Jordan	530 Netherlands Antilles
404 Kenya	536 Neutral Zone
408 Korea, Democratic People's Republic of	540 New Caledonia
410 Korea, Republic of	554 New Zealand
414 Kuwait	558 Nicaragua
418 Lao People's Democratic Republic	562 Niger
422 Lebanon	566 Nigeria
426 Lesotho	570 Niue
430 Liberia	574 Norfolk Island
434 Libyan Arab Jamahiriya	578 Norway
438 Liechtenstein	512 Oman
442 Luxembourg	582 Pacific Islands (Trust Territory)
446 Macau	586 Pakistan
450 Madagascar	591 Panama
454 Malawi	598 Papua New Guinea
458 Malaysia	600 Paraguay
462 Maldives	604 Peru
466 Mali	608 Philippines
470 Malta	612 Pitcairn Island
474 Martinique	616 Poland
478 Mauritania	620 Portugal
480 Mauritius	630 Puerto Rico
484 Mexico	634 Qatar
488 Midway Islands	638 Reunion
492 Monaco	642 Romania
	646 Rwanda
	650 Ryukyu Islands

732 Saharan Arab Democratic Republic	772 Tokelau Islands
654 St Helena	776 Tonga
658 St Kitts-Nevis-Anguilla	780 Trinidad and Tobago
662 St Lucia	788 Tunisia
666 St Pierre and Miquelon	792 Turkey
670 St Vincent	796 Turks and Caicos Islands
674 San Marino	800 Uganda
678 Sao Tom and Principe	804 Ukrainian Soviet Socialist Republic
682 Saudi Arabia	810 Union of Soviet Socialist Republics
686 Senegal	784 United Arab Emirates
690 Seychelles	826 United Kingdom
694 Sierra Leone	834 United Republic of Tanzania
698 Sikkim	840 United States of America
702 Singapore	858 Uruguay
090 Solomon Islands	548 Vanuatu
706 Somalia	862 Venezuela
710 South Africa	704 Viet Nam
724 Spain	850 Virgin Islands
728 Spanish North Africa	872 Wake Islands
144 Sri Lanka	876 Wallis and Futuna Islands
736 Sudan	882 Western Samoa
740 Suriname	886 Yemen
744 Svalbard and Jan Mayen Islands	890 Yugoslavia
748 Swaziland	180 Zaire
752 Sweden	894 Zambia
756 Switzerland	716 Zimbabwe
760 Syrian Arab Republic	896 Areas not elsewhere specified
158 Taiwan	898 Not specified
764 Thailand	900+ special codes for areas
626 Timor	etc. required by user
768 Togo	

引用文献: Department of Economic and Social Affairs, Statistical office of the United Nations, Statistical papers, Series M No. 49, United Nations, New York, 1970

付録 2. データ項目の整合性に関する編集作業

V. van Holten

データ項目の整合性に関する編集は、コーディングの前、または後に実行される。キー入力するときにデータを編集する、いわゆるインテリジェント・データ入力端末の利用が次第に一般化されつつある。このプロセスによって、データが正しくコード化されたかどうかチェックされ、さらに入力したときに、フィールド間で整合性がチェックされる。また編集作業の長所はデータが永久データベースに追加される前に、データの整合性のチェックと修正が実行されるという長所がある。しかしながらこの手法では、キー入力者が、問題を解決する能力を持っているか、もしくは、問題が他の誰かによって解決されるまでレコードを保持しておく必要がある。編集や整合性チェックは、すべてのデータが入力された後に実行してもよい（バッチ様式では）。さらに、新しく入ってくるデータと、既に登録室に報告されているデータとの間で整合性をチェックするためにも、コンピュータを用いることができる。整合性の不一致が発見されたら、登録室のスタッフが問題を解決できるようにリストが作成されるか、あるいは、既に提出されたデータを、人手を介することなく、コンピュータによって更新するアルゴリズムを確立することができる。

登録室の最も重要な機能の一つは、ある患者に関する種々の情報を統合することである。それ故、蓄積されたデータが内部で整合性を保つように、あらゆる努力がなされなければならない。ここに示す例は United States Surveillance, Epidemiology, and End Results (米国 SEER) Programに提出された1レコードにおけるデータ項目間、及び複数レコード間の整合性チェックの具体例である。SEER Programは、13の地域がん登録の共同体であり、各登録は、毎年データをコード化した形で US National Cancer Institute (米国 NCI)に報告する。NCI はデータをコード化した形でのみ受け取るので、発見された問題のすべては、解決のために再び個々の地域がん登録に問い合わせなければならない。個々の独立の原発性腫瘍について、1つのデータレコードが NCI に提出される。しかし、独立した2個以上の原発腫瘍を持つ症例については、患者登録番号は、各々のデータレコードで同一である。従ってレコード間でも整合性を同様にチェックすることができる。

以下に、現在 SEER Program で利用されている50の編集作業のいくつかを表示する。表示した編集作業は、どの登録室でも有益であると考えられるので、取り上げた。これらは、殆んどのがん登録室で収集しているデータフィールドに当てはまる。SEER Program で行われるその他の編集作業は、SEER のデータベースに特異的なフィールド（例えば登録室 I D）か、あるいは SEER がデータ収集を開始して以来、変更を加えた報告要求項目に適用される。

これらの編集の中には、通常予期されるコードの組み合わせは正しいと考え、通常予期しないコードの組み合わせは誤りと考えるようにデザインされているものがある。それ故、正

しいが稀なフィールドの組み合わせが、誤りとしてマークされることがあり得る。これらの稀な症例の情報を保持するため、編集用にオーバーライド・フラッグという別のフィールドが創設される。オーバーライド・フラッグが「オン」にセットされると、もはやその症例には不一致がないと見なされる。例えば年齢と原発部位の編集において、18歳の子宮頸部浸潤がんという稀な症例について、オーバーライド・フラッグを「オン」にセットすることができる。しかし、このような症例については、将来、レビューが必要となる。

SEER 計画におけるフィールド間編集

SEER programで用いられるフィールド間編集のいくつかを、表1に示す。

表1. SEERにおけるフィールド間編集(フィールドAとフィールドBの組み合わせ)

(注: 下記の編集には ICD-0 1976年版を引用)

フィールド A 項目名 ¹	フィールド B 項目名 ¹	編集規準
資料源(項目35)	予後(項目31)	もし剖検または死亡票のみのケースであれば、予後調査結果は死亡でなければならない。
資料源(項目35)	死因(項目33)	もし剖検または死亡票のみのケースであれば、死因が特定されていなければならない。
罹患時年齢(項目9)	婚姻状態(項目8)	15歳未満の場合には婚姻状態が未婚でなければならない。
罹患時年齢(項目9)	出生日(年/月) (項目5)と罹患日 (年/月)(項目16)	年齢は計算年齢 [((罹患年×12+罹患月)-(出生年×12+出生月))/12]と一致しなければならない。
出生日(項目5)	罹患日(項目16)	罹患日は出生日と同じか大きくななければならない。
性(項目4)	原発部位(項目20)	原発部位コードが女性乳房(174.-)、女性生殖器(179.9184.9)は、男にはありえない。 原発部位コードが男性乳房(175.9)、男性生殖器(185.9187.9)は、女にはありえない。
罹患時年齢(項目9)	原発部位(項目20)と 組織型(項目21)	もしも、ケースが既にレビューされたことを示すオーバーライドフラッグがオンであれば、編集に関連する他のフィールドに変更が加えられてない限り、これ以上のチェックは実行されない。

表1-続き

(注: 下記の編集には ICD-0 1976年版を引用)

フィールド A 項目名 ¹	フィールド B 項目名 ¹	編集規準
	年齢/組織型の編集 に関するオパールライト フラッグ	もしも5歳未満なら、下記の原発部位はありえない 子宮頸(180.-) 前立腺(185.9)
		もしも20歳未満なら、以下の原発部位はありえない 食道(150.-) 小腸(152.-) 結腸(153.-); (組織型はカルチノイド(M8240- 8244)でない 直腸、直腸S状結腸移行部、肛門管と肛門、 NOS(154.-) 胆嚢及び肝外胆管(156.-) 膵(157.-) 気管、気管支及び肺(162.-)もし組織型がカルチ ノイド(M8240-8244)でなければ 胸膜(163.-) 乳房(174.-, 175.9) 子宮 NOS(179.9) 浸潤性の性状コードをもつ子宮頸(180.-) 子宮体(182.)
		もしも30歳未満なら、以下の組織型はありえない。 形質細胞骨髄腫(M9730) 慢性リンパ性白血病(M9823) 慢性骨髄性白血病(M9863) 単球性白血病、NOS(M9890)
		もしも45歳未満なら、原発部位は腺癌(M8140)の組 織型をもつ前立腺(185.9)ではありえない。
		もしも5歳以上の場合は、原発部位が網膜芽腫 (M9510-9512)の組織型を有する眼(190.-)ではあり 得ない。
		もしも15歳未満か45歳を越えるなら、原発部位は 絨毛癌(M9100)の組織型を有する胎盤(181.9)では あり得ない。
罹患日(項目16)	がん治療開始日	罹患日は、がん治療開始日と同じか、以前でなけ ればならない。

表1 - 続き

(注: 下記の編集は ICD-0 1976年版を引用)

フィールド A 項目名 ¹	フィールド B 項目名 ¹	編集規準
罹患日(項目16)	最終接触日(項目30)	罹患日は、最終接触日か、死亡日と同じ、または以前でなければならない。
原発部位(項目20)	組織型(項目21)	この編集は、部位/組織型の妥当性編集の記述で示されている。
原発部位(項目20) と組織型(項目21)	病変範囲 (項目23-25)	この編集は、各部位と組織型の組み合わせに対して許容された病変範囲コードに従って実行される。
原発部位(項目20)	側性(左右の別) (項目28)	以下の ICD-0 部位は、正当な側性(左右の別)コードを持っていないなければならない: 142.0 耳下腺 142.1 顎下腺 142.2 舌下腺 146.0 扁桃 146.1 扁桃窩 146.2 口蓋弓 160.1 中耳 160.2 上顎洞 160.4 前頭洞 162.3 上葉、肺 162.4 中葉、肺 162.5 下葉、肺 162.8 肺 162.9 肺、NOS 163.- 胸膜 170.4 上肢の長骨、肩甲骨及びその関節 170.5 上肢の短骨及びその関節 170.7 下肢の長骨及びその関節 170.8 下肢の短骨及びその関節 171.2 上肢、肩の結合組織、皮下組織及びその他の軟部組織 171.3 下肢、股関節部の結合組織、皮下組織及びその他の軟部組織 173.1 眼瞼 173.2 外耳 173.3 その他の部位及び部位不明の顔面の皮膚 173.5 体幹の皮膚 173.6 上肢及び肩の皮膚

表1-続き

(注: 下記の編集には ICD-0 1976年版を引用)

フィールド A 項目名 ¹	フィールド B 項目名 ¹	編集規準
原発部位(項目20)	側性(左右の別) (項目28)	以下の ICD-0 部位は、正当な側性(左右の別)コードを持っていないなければならない: 173.7 下肢及び股関節部の皮膚 174.- 女性の乳房 175.9 男性の乳房、NOS 183.0 卵巣 183.2 卵管 186.0 停留睪丸 186.9 睪丸、NOS 187.5 副睪丸(精巣上皮) 187.6 精索 189.0 腎、NOS 189.2 尿管 190.- 眼及び涙器 194.0 副腎 194.5 頸動脈小体
形態(項目22)	診断根拠(項目17)	性状コードが上皮内がんの場合、組織診による確証がある。
形態(項目22)	疾患の進展 (項目23-25)	性状コードが上皮内がんの場合、疾患の進展度は上皮内でなければならない。
がん治療開始日	がん治療の初回コース (項目29)	治療日は、治療の初回コースが治療実施を示していたならば、正当な日付を持っていないならない。
追跡状況(項目31)	死因(項目33)	追跡状況が生存の場合、死因は 0000 (生存) でなければならない。追跡状況が死亡の場合、死因は0000であってはならない。
最終接触日	がん治療開始日 (項目30)	がん治療開始日は、最終接触日と同じか以前でなければならない。

¹ ()内の項目番号は、第6章で述べられている患者情報の項目番号と一致する。

SEER計画における部位と組織型との組合せの妥当性についての編集

部位と組織型（性状コードを含む）の双方が、それぞれ（性質コードを含めて）信頼しうることが判明したならば、次に部位／組織型の組み合わせに関する編集が実行される。部位／組織型の組み合わせにおける妥当性は、部位毎に与えた許容されうる組織型コードかどうかによりチェックされる。部位／組織型の妥当性リストは、各部位につき妥当と考える各々の部位と組織型コード（4桁数字）を明示する。例えば口唇がんの場合を表2に示す。このリストは、しばしばある範囲の部位コードを特定しているが、その範囲内の部位コードだけが適用となる。部位／組織型の妥当性リストでの組織型用語は、必ずしも腫瘍学のための国際疾病分類(ICD-0)にある好ましい用語ではない。組織型コードが、部位と照らして妥当であると明示されていない症例については、要再検査の信号が出力される。もし、そのケースを再検討した後に、部位／組織型の組み合わせが正しいとわかれば、当然オーバーライドフラッグを「オン」にセットすることになる。

なお、以下の組み合わせは、妥当性がない。すなわち、

(1) 部位不明あるいは不明確な部位と上皮内がんの性状コードを有する組織型：

- 149.9 部位不明確の口唇、口腔及び咽頭
- 159.9 胃腸管、NOS
- 165.9 呼吸系、部位不明確
- 179.9 子宮、NOS
- 184.9 女性生殖器、NOS
- 187.9 男性生殖器、NOS
- 189.9 泌尿系、NOS
- 192.9 神経系、NOS
- 194.9 内分泌腺、NOS
- 195. その他の不明確な部位
- 199.9 原発部位不明

(2) 不明確な部位(195.)と黒色腫(M8720-8790)を示す組織型コード

表2. SEER計画における部位と組織型との組合せの妥当性についてのリストの例

部位：口唇 1400-1409 について。以下の組織型は整合する。

組織型(4桁数字)	組織型 (4桁数字)
8000/3 新生物、悪性	8081/2 ボーエン病
8001/3 腫瘍細胞、悪性	8082/3 リンパ上皮癌
8002/3 悪性腫瘍、小細胞型	
8003/3 悪性腫瘍、巨細胞型	8140/2 上皮内腺癌
8004/3 悪性腫瘍、紡錘形細胞型	8140/3 腺癌、NOS
	8141/3 硬性腺癌
8010/2 上皮内癌、NOS	8143/3 表層拡大性腺癌
8010/3 癌、NOS	
8011/3 上皮腫、悪性	8200/3 腺様嚢胞癌
8012/3 大細胞癌、NOS	8201/3 篩状癌
8020/3 癌、未分化型、NOS	8260/3 乳頭状腺癌、NOS
8021/3 癌、退形成型、NOS	8261/2 絨毛状腺腫内上皮内腺癌
8022/3 多形細胞性癌	8261/3 絨毛状腺腫内腺癌
	8262/3 絨毛状腺癌
8030/3 巨細胞及び紡錘形細胞癌	
8031/3 巨細胞癌	8263/2 管状絨毛状腺腫内上皮内腺癌
8032/3 紡錘形細胞癌	8263/3 管状絨毛状腺腫内腺癌
8033/3 偽肉腫様癌	8430/3 粘表皮癌
8034/3 多角形細胞癌	8480/3 粘液性腺癌
8050/2 乳頭状上皮内癌	8481/3 粘液産性腺癌
8050/3 乳頭状癌、NOS	8720/3 悪性黒色腫、NOS
8051/3 疣状癌、NOS	8721/3 結節性黒色腫
8052/3 乳頭状扁平上皮癌	8722/3 風船細胞黒色腫
8070/2 上皮内扁平上皮癌、NOS	8730/3 無色素性黒色腫
8070/3 扁平上皮癌、NOS	8743/3 表層拡大性黒色腫
8071/3 扁平上皮癌、角化型、NOS	8771/3 類上皮細胞黒色腫
8072/3 扁平上皮癌、大細胞性、非角化型	8772/3 紡錘形細胞黒色腫、NOS
8073/3 扁平上皮癌、小細胞性、非角化型	8775/3 類上皮細胞と
8074/3 扁平上皮癌、紡錘形細胞型	紡錘形細胞混合性黒色腫
8075/3 腺様扁平上皮癌	8940/3 混合腫瘍、悪性、NOS
8076/2 上皮内扁平上皮癌	8941/3 多形性腺腫内がん
8076/3 扁平上皮癌、微小侵襲性	9140/3 カポジー肉腫

SEER計画におけるレコード間の編集

1人の患者がファイルに複数のレコードを持つ場合は、常に、レコード間での適切な整合性を確認するため、以下のフィールドについて編集が実行される。

出生地（項目 7）

すべてのレコードは同じでなければならない。

出生年月日（項目 5）

すべてのレコードは同じでなければならない。

人種（項目 11）

すべてのレコードは同じでなければならない。

性（項目 4）

すべてのレコードは同じでなければならない。

一連番号（腫瘍同定）

- (1) 1人の患者について複数のレコードがある場合、一連番号に0あるいは不明のレコードがあってはならない。
- (2) 一連番号には、同じ番号があってはならない。

一連番号／罹患日

腫瘍の一連番号は、原発腫瘍の罹患の時間的順序を反映するものでなければならない。従って、一連番号に1を割り当てられた原発腫瘍の罹患日は、一連番号に2を与えられた原発腫瘍の罹患日に比べ、同日、あるいはそれ以前でなければならない。

予後調査日あるいは死亡日

すべてのレコードにあるこれらの日付は、同一でなければならない。

予後調査状況（項目 31）

- (1) もしどれかのレコードで患者が生存と示していれば、その他のレコードも生存と明示していなければならない。
- (2) もしどれかのレコードで死亡と示していれば、その他のレコードも死亡と明示していなければならない。

表3. 再点検が必要な原発部位の組み合わせ

不明確または NOS の部位	特定部位
149.9 口唇、口腔及び咽頭の不明確な部位	140.0-149.8
159.0 腸管、NOS	150.0-158.9
159.8 消化器系のまたがる部位	150.0-158.9
159.9 消化器系及び腹膜の不明確な部位	150.0-158.9
165.0 上気道	160.0-163.9
165.8 呼吸器及び胸腔内のまたがる部位	160.0-164.9
165.9 呼吸系の不明確な部位	160.0-163.9
184.9 女性生殖器、NOS	179.9-184.8
187.9 男性生殖器、NOS	185.9-187.8
189.9 泌尿系、NOS	188.0-189.8
194.8 内分泌腺多発性	193.9-194.6
194.9 内分泌腺、NOS	193.9-194.8
1AA.8a いずれかの overlapping 部位コード	1AA.xa, b いずれかの詳細部位
1BB.9c いずれかの NOS 部位	1BB.xb, c いずれかの詳細部位

- a : AA は、46, 51, 54, 58, 62, 70, 74, 80, 83, 87, 91を除く40-99までのいずれかの2桁数字
 b : x は、いずれかの1桁数字
 c : BBは、40-99までのいずれかの2桁数字

多重がんを二重登録から区別すること

互いに独立した多重がんであると報告されてきたが、本当は1腫瘍であった例をチェックする目的で、編集プログラムを作成することができる。この種の編集は、浸潤がん症例にのみ適用され、そのフォーマットは「多重がん」をきめるために用いた定義によって決定される。もし第7章の定義が使われたのであれば、その2症例は、同一組織型群（第7章、表2の1群、2群、3群、5群、6群、もしくは7群）に属するものでなければならない。

1つの部位が不明確もしくは NOS（詳細部位不明）の部位で、いま1つの部位が詳細部位をもつ場合には、そのレコードには、再検討のための印をつける。もし、再検討の後、ケースが多重がんとして決定されれば、その患者の原発部位が再度検討され、正しいことがわかったことを指し示すために、オーバーライド・フラッグが用いられる。表3は、再検討が必要な原発部位の組み合わせを示している。

（翻訳：藤田委由、津熊秀明）

付録3 (a) デンマークがん登録

能動的なデータ収集の要素を兼ね備えた 篤志届出による全国がん登録システム

Hans H. Storm

緒言

デンマークの保健事業

デンマーク王国（グリーンランドとフェロエを除く）は北緯55° から58°、東経8° から12.5° の間にあり、面積は43,080km²である。1986年1月1日現在の人口は510万人である。保健医療システムは、国民健康保険と契約を結んだ開業医と専門医からなる私的セクターと、公立、公的、および国立の病院から成る公的セクターとで構成されている。医療は全居住者に対し無料で提供される。

1980年の病院病床数は人口千対5.6床、医師数は人口千対2.2名である。がんの外科的治療は、一般病院とがんセンターの双方で実施される。病院部門は、28の病理検査施設のサービスを受けている。非外科的治療をうける患者は、5つの地域放射線治療・がんセンターに、多く集中している。殆んど全ての住民は、数時間以内に地域がんセンターの1つを受診することができる。

デンマークがん登録事業

デンマークがん登録は、デンマーク国民の間に発生した全がん患者を登録するために、全国規模のプログラムとして1942年5月に設立された。デンマークがん登録は、国家保健会議を代表して、デンマーク対がん協会が運営しており、デンマーク医師会の支援を受けている。罹患率は1943年1月1日から計算されている。

がん登録の元来の使命（Clemmesen, 1965）は、以下の諸統計を提供しうるようにするため、基礎資料を収集することである。

- (a) がんの治療結果を正確に評価するための信頼性の高い罹患統計。
- (b) 時代、地域、職業などの差によるがん罹患の相違の正確な測定。
- (c) 重複がん、同時発生がん等の研究のみならず、臨床医に役立つような個々の患者に関する統計。

がん登録は、時代、地理的位置、職業、その他の要因について、がん罹患の変移をモニターするという展望をもって、罹患統計を作成する。また、がんの原因と予防の分野で疫学研究を実施する。

罹患の基本資料

登録は腫瘍を指向しており、腫瘍が登録の単位である（いくつかの例外がある。すなわち類似の形態的特徴を有する皮膚や一対臓器の重複がん）。登録事業の最初の25年間は、ある人の特定の腫瘍記録を結びつけるキーとなる同定項目は、誕生日と氏名であった。1968年以降は、現在キー同定項目として使われている特異的個人同定番号（PNR）の導入によって、登録事業が促進された。この同定番号は、内部のリンケイジを促進し、重複登録を防ぐ。

医療機関からの届出

法的側面

1987年3月まで、医療機関からの届出は、篤志により行われ、受け取った各届出様式に対して少額の金銭が支払われていた。1987年3月1日に、がん患者の届出が義務づけられたが、（Sundhedsstyrelsen, 1987）その他の登録業務には、1943年以降、届出システムに変更はない。法的には、がん登録室への届出の責任は、病院の各部長、剖検を実行する病理部長、あるいは、患者を病院に紹介せずに、自ら患者の治療とフォローアップを実施する開業医、などが負う。

届出すべき疾患と届出

① がん、肉腫、白血病、リンパ腫を含むすべての悪性腫瘍と、② すべての脳及び中枢神経系腫瘍と、③ 性状に関係なく全膀胱腫瘍と、④ 子宮頸部のすべての前がん病変とは、デンマークがん登録室に届出しなければならない。その他の前がん性病変は届出の必要がない。届出の必要な疾患は、以下のICD-8分類（本分類はデンマーク保健省によって現在も用いられている）に該当する：ICD-8 140-207, 223, 225, 230-239。

届出規則によれば、届出の必要な疾患全ての新規診断例を届出すべきこと、患者が重複がんを有している場合は、各々の原発腫瘍についてがん登録室に別々の報告書を提出しなければならないこと、が規定されている。以下の諸点も届出なければならない。

- 既に届出たケースの診断に関するあらゆる改訂；
- 異型性病変の進展；
- 前がん性病変、もしくは上皮内がんの進展；
- 新規に、もしくはこれまでに診断されたがん症例の剖検結果。

データ源

デンマークがん登録は、表1に示すように、死亡票に加えて、臨床病院部、病理部と開業医より届出を受ける。以下で詳細を述べる。この情報は、中央人口登録に由来する個人情報によって補足される。

表1. デンマークがん登録のデータ源

医学分野：	
届出票：	開業医（一般医） 臨床専門医 病院部門 病理検査施設 法医学施設
死亡診断書：	コンピューター化したデータとマイクロフィルム
非医学分野：	
中央人口登録局：	（コンピューター化）
地方人口登録局：	（非コンピューター化）

医学データ源

病院各部は、治療、診断のために新規に入院したあらゆるがん患者を、がん登録室に届出するように要請されている。典型的な場合、患者は地方レベルの病院医療施設を受診し、その後、より高度な診断、治療のため、より専門の各部門に紹介される。多くのがん患者は、放射線治療や高度の専門的治療のために、デンマークにある5つの主要がんセンターの1つに紹介される。がん登録室は、それ故、一人のがん患者について複数の届出を受け取る。例え1、2の病院部門があるケースを届出しなかったとしても、がん登録がその他の部門からの届出により、そのケースを把握する機会がかなりある。デンマークがん登録事業の高度の完全性は、このように、デンマークの保健医療システムの結果である(Osterlind & Jensen, 1985; Storm, 1988)。

臨床病院部門は、臨床届出票を記入し、提出するように求められており、診断X線検査や組織病理検査などの特殊サービス部門由来の成績をも記載する。治療やフォローを担当する施設からの情報は、病理や診断X線検査部門のような、その他の施設に由来するデータのコーディングを容易にし、誤った解釈を回避するのに役立つ。開業医は、それまで以上の治療や診断のために、病院へ紹介されることがなかったがん患者を届出するように要請されている。

デンマークの剖検率は高く(35%)、全がん死亡の43%が剖検される(Storm & Andersen, 1986)。がん患者の剖検記録は、腫瘍組織の有無にかかわらず、病理施設から登録室に直接報告される。死亡前にはがんを疑わず剖検でのみがんと診断されたケースも報告され、登録に加えられる。1943年以来、デンマークがん登録室は、死亡診断書に原因死もしくは寄与死因と記載されたケースの情報を得てきた。そうしたケースの同定は、国の死亡診断書システム

との手作業によるリンケイジにより実施されてきたが、1971年からは、リンケイジがコンピュータ化された。死亡票のみによって確認されたがん患者の割合は、表2に示すように、1943-47年の約19%から、1977-82年の1-2%に減少した(Jensen et al., 1985)。デンマーク登録室は、こうした死亡票のみの症例を再調査し、死亡票に記入したがん登録室に届出しなかった開業医もしくは病院部門の医師に届出することを要求する。

非医学的データ源

登録室で受け入れられるがん患者はすべて、診断日においてデンマークの居住者でなければならない。中央人口登録局 (CPR) では、デンマークの全居住者についてのファイルを常に更新しており、これには、氏名、住所、婚姻状況、国外への転出と転入、職業、及び死亡について情報を備えている。更新の日付に関する詳細、また婚姻状態や住所等の変化に関する履歴情報を得ることもできる。全ての届出がん患者は、出生時あるいは永久にデンマークでの居住を定めたときに割り当てられる特異的な個人同定番号 (PNR) を使って、CPRとリンケイジされる。このリンケイジには2つの目的がある。すなわち、診断日を参照しながら、届出された人のアイデンティティーをチェックすること、及び上述の項目に関する名前等の情報を移入することである。

将来のデータ源

保健医療システムの中で、現在コンピュータ化が広がっている。全国病院退院患者登録システム (HDR) は、1970年代後期より稼働している。HDRは、将来、届出のなかったがん患者の同定に利用できるかもしれない(Osterlind & Jensen, 1985)。SNOMED 分類を用いた病理診断のコンピュータシステムが、現在デンマーク病理部の50%で稼働している。このシステムもがん患者の同定に使える可能性があり、届出内容の不十分ながん患者の形態学に関する情報を、恐らく提供することができる。これらのシステムは、いずれもがん登録のために創設されたものではなく、通常の登録制度を単に補足するだけかもしれない。これらのシステムの持つ主要な問題点は、診断 (あらゆる非がん診断を含む) の中央化されない解釈とコーディングであり、これらはしばしば経験の少ない医療スタッフや非医療スタッフにより行われており、その結果、診断情報の不正確さを導く可能性がある。

届出票

使われている届出票は単純で、各ケースについて限られた量の情報を求めている。届出票は、概ね自明で、自己記入式である。氏名、職業、これまでの治療病院、腫瘍と治療の詳細、及び種々の出来事の日付を入れるために設けたスペースとともに、特定の質問をチェックするボックスがある。届出票は複写コピー用紙でできており、届出票コピーは病院記録に保管される。届出票に記入するための簡単な説明が、届出票コピーの裏面にある。これらの説明

表2. がん患者の診断根拠別の百分率分布

診断年別, 全がん, 男女計, デンマーク, 1943-80

診断年	登録数	診断根拠 (%)				
		顕微鏡で確認		顕微鏡 確認なし		
		剖検あり	剖検なし	臨床診断のみ	剖検のみ	死亡票のみ
1943-47	49,144	37.1	15.6	23.5	4.6	19.3
1948-52	55,439	43.3	19.6	20.8	2.6	13.7
1953-57	64,238	43.8	24.3	18.0	2.3	11.7
1958-62	74,043	46.6	29.0	14.4	2.0	8.0
1963-67	86,902	48.9	32.0	11.0	2.0	6.1
1968-72	103,282	50.5	35.2	7.9	1.9	4.5
1973-77	118,118	54.6	31.8	6.9	1.8	5.0
1978-80	79,921	65.6	24.8	7.6	0.7	1.3
全期間	631,087	50.2	28.0	12.2	2.1	7.5

は、届出を行う全ての施設に郵送される詳細な説明のある小冊子によって、補完されている。2種類の様式が利用されており、1つは開業医と病院部門用(図1)で、いま1つは病理施設用(図2)である。

図1. 臨床部門で使用する届出票の裏側

図2. 病理部門で使用する届出票の裏側

訳者注：表側が届出票、裏側が届出票記載要領になっているが、表側はデンマーク語で書かれており、翻訳できない部分が多かったので、省略し、英語で書かれた裏側のみを、次々頁とその次の頁とに掲載した。

臨床届出票(次々頁に裏側のみ転載)

腫瘍診断は、部位及び形態ともに、明快地記述するよう、要求される。診断日には、悪性疾患についての初回入院年月、もしくは外来受診年月日を採用する。疾患の進展度は、前がん病変、限局、領域転移、遠隔転移あるいは不明と、あらかじめ分類したボックスにチェックする。婦人科腫瘍に対する FIGO のように、ある種の腫瘍については、より詳細な病期分

類が、明確な記述で示されることがある。診断根拠は、組織診、骨髄検査、細胞診、組織診のない外科手術、その他の特異検査、臨床診断のみ、及び剖検による偶発的診断、に分けられたボックスをチェックすることにより示される。

治療の情報は乏しく、初期治療、すなわち、最初の4ヶ月間に行われた治療のみが記録される。外科治療は、「はい」と「いいえ」をチェックし、手術日を記入し、明確な言葉で手術方法、例えば「結腸切除術」と記載する。医師は、外科処置が単に診断目的であったのか、姑息的もしくは根治を目指したものであったかを示すように要請される。その他の治療については、放射線治療、化学（細胞毒性）療法、ホルモン療法について記入されるが、詳細は省かれる。治療開始日の記入が求められる。なお、治療なし、あるいはその他の治療をチェックし、この点を明確に記述することも可能である。ボックスのチェックは、届出票に記述された情報（組織診、外科手術や剖検成績など）とクロスチェックされる。死亡患者については、死亡日が当然記述される。もしも剖検が実施されたならば、剖検を実施した病院と部門（の名称）が、がんに関する総括的結論に加えて、記入される。さらに、その剖検を施行した病理医は、病理施設用の特別の届出票で所見を報告するように要請される。

病理（剖検）届出票（次々頁に裏側のみ転載）

病理部門からの剖検所見に関する届出票は、臨床部門からの届出票と、治療を除いて同じ情報を求めている。病理医は、患者を治療した部門の名称を記入するよう要請されている。これにより、登録室は、もしもそのケースがこれまでに届出されていない場合には、患者を治療した臨床部門に届出票を要求することができる。数種のがんが剖検時に発見されることがあるので、病理届出票では、1人につき3個のがんが記入できるスペースを設けている。

登録方法

届出票の受理

届出票は毎日郵送され、そして、図3で示す流れ図に従い、週毎にバッチ処理される（約1,000票）。初期相は、第1段階として、カレンダー・ウィークと名付けるデータベースの作成を含んでいる。届出票が届くと、個人同定番号、届出病院の同定番号、及び到着の週がデータベースに入力される。個人同定番号と届出病院同定番号のコンピュータによる妥当性チェックが実施される。登録室のスタッフが目でみて、届出票の内容チェックを行ない、もしも重要な欠落がある場合には届出医師に連絡する。届出票の週毎のバッチ処理においては、登録室において同じ患者について既届出があるかどうかのコンピュータによる検索が行われ、その後、既登録患者と、新患者とに区分される。既登録のケースのデータは既にマイクロフィルムに存在するので、新旧2群の患者情報のバッチ処理は幾分異なる。ある個人についての既存の全情報（マイクロフィルム、処理中の様式、及びコンピュータ化された情報）が呼び

REVERSE SIDE OF NOTIFICATION FORM CONCERNING MALIGNANT OR PREMALIGNANT TUMOURS

Notification to the Cancer Registry is compulsory for all physicians as written in instruction no. 50 from the Danish Board of Health of 15 January 1987: "Instruction for physicians concerning notification to the Cancer Registry of cases of cancer":

INSTRUCTIONS

WHICH DISEASES ARE TO BE REPORTED?

All cases of malignant tumours, such as carcinomas, leukaemias, malignant lymphomas and all brain tumours (including benign), all bladder tumours and uterine precancerous lesions are to be reported to the Cancer Registry.

Previously diagnosed tumours that are not refound at autopsy should be indicated under the heading "remarks"; e.g. Neoplasma malignum pulmonum dextri (adenocarc.) tractum. 1978—not found at autopsy.

All cases of doubt should be reported.

NOTIFICATION SHOULD TAKE PLACE FOR:

1. All cases of newly diagnosed tumours.
2. Cases of multiple tumours in the same person, with separate reports for each tumour that is considered a new primary tumour.
3. Revision of previous diagnosis.
4. The ascertainment that a previously reported tumour did not exist after all.
5. The progression of precancerous lesions or carcinoma in situ, including a change to an invasive tumour.

The report should be submitted, at the latest, when the patient is discharged from the hospital.

OCCUPATION

The patient's trade or profession should be specified. Please avoid imprecise statements; e.g. "bank manager" should be stated instead of just "manager", "journeyman carpenter" instead of just "carpenter", and "farm owner", instead of just "farmer". For retired people please state their former occupation also, e.g. "former bricklayer".

HOSPITAL AND DEPARTMENT OR GENERAL PRACTITIONER

The name of the department and hospital reporting should be stated. For reports from a general practitioner or specialist, the name and address of the physician should be stated.

DIAGNOSIS

State the discharge diagnosis, and describe the exact position of the primary tumour, as regards both organ and localization of the tumour in the organ.

THE HISTOLOGICAL DIAGNOSIS OF THE TUMOUR should be stated.

If a histological examination has only been carried out for metastases, please state. For cancer in the bladder and papillomas, state a grade from I to IV.

Stages for cancer of the cervix uteri should be stated. The spreading of the tumour should be ticked.

TREATMENT

Treatment directed at the primary tumour or metastases during the present admission should be stated.

DEATH

If the patient is deceased please give the information required.

AUTOPSY

Autopsy number, hospital where autopsied, and tumour diagnosis should be stated (please see detailed instruction).

INCIDENTAL AUTOPSY FINDING

For every tumour, tick whether the tumour was an incidental autopsy finding, i.e. a disease (tumour) which gave no symptom or objective sign when the patient was alive.

INSTRUCTIONS

WHICH DISEASES ARE TO BE REPORTED?

All cases of malignant tumours, such as carcinomas, leukaemias, malignant lymphomas and all brain tumours (including benign), all bladder tumours and uterine precancerous lesions are to be reported to the Cancer Registry.

Previously diagnosed tumours that are not refound at autopsy should be indicated under the heading "remarks"; e.g. Neoplasma malignum pulmonis dextri (adenocarc.) tractatum. 1978—not found at autopsy.

INSTRUCTIONS FOR FILLING IN THE NOTIFICATION FORM FOR THE CANCER REGISTRY REGARDING AUTOPSIED CASES OF MALIGNANT DISEASE:

1. OCCUPATION

The patient's trade or profession should be specified. Please avoid imprecise statements; e.g. "bank manager" should be stated instead of just "manager", "journeyman carpenter" instead of just "carpenter", and "farm owner", instead of just "farmer". For retired people please state their former occupation also, e.g. "former bricklayer".

2. ADMISSION

State the date of admission and name of the hospital and department. Date of death should be stated for a person who has not been in hospital.

3. AUTOPSY

Autopsy number and hospital should be stated. Reference number for departments according to the Danish Board of Health's Classification of Hospitals should be stated (valid as per 1 January 1982).

4. AUTOPSY DIAGNOSIS AND MORPHOLOGICAL DIAGNOSIS

For every single tumour, even unverified, the final autopsy diagnosis should be stated, including information on the anatomical localization and morphology.

N.B. If the primary tumour is unknown, this must be evident from the diagnosis. The expression "cancer metastaticus" can be used, e.g. c. metast. pulm. dext. or c. metast. hepatitis. For every gynaecological tumour the stage should be indicated. For cancer of the bladder (including papilloma), grade I to IV should be stated.

It should be indicated whether there has been histological verification and whether there is spreading of the tumour.

5. INCIDENTAL AUTOPSY FINDING

For every tumour, tick whether the tumour was an incidental (unexpected) autopsy finding, i.e. a disease (tumour) which gave no symptom or objective sign when the patient was alive.

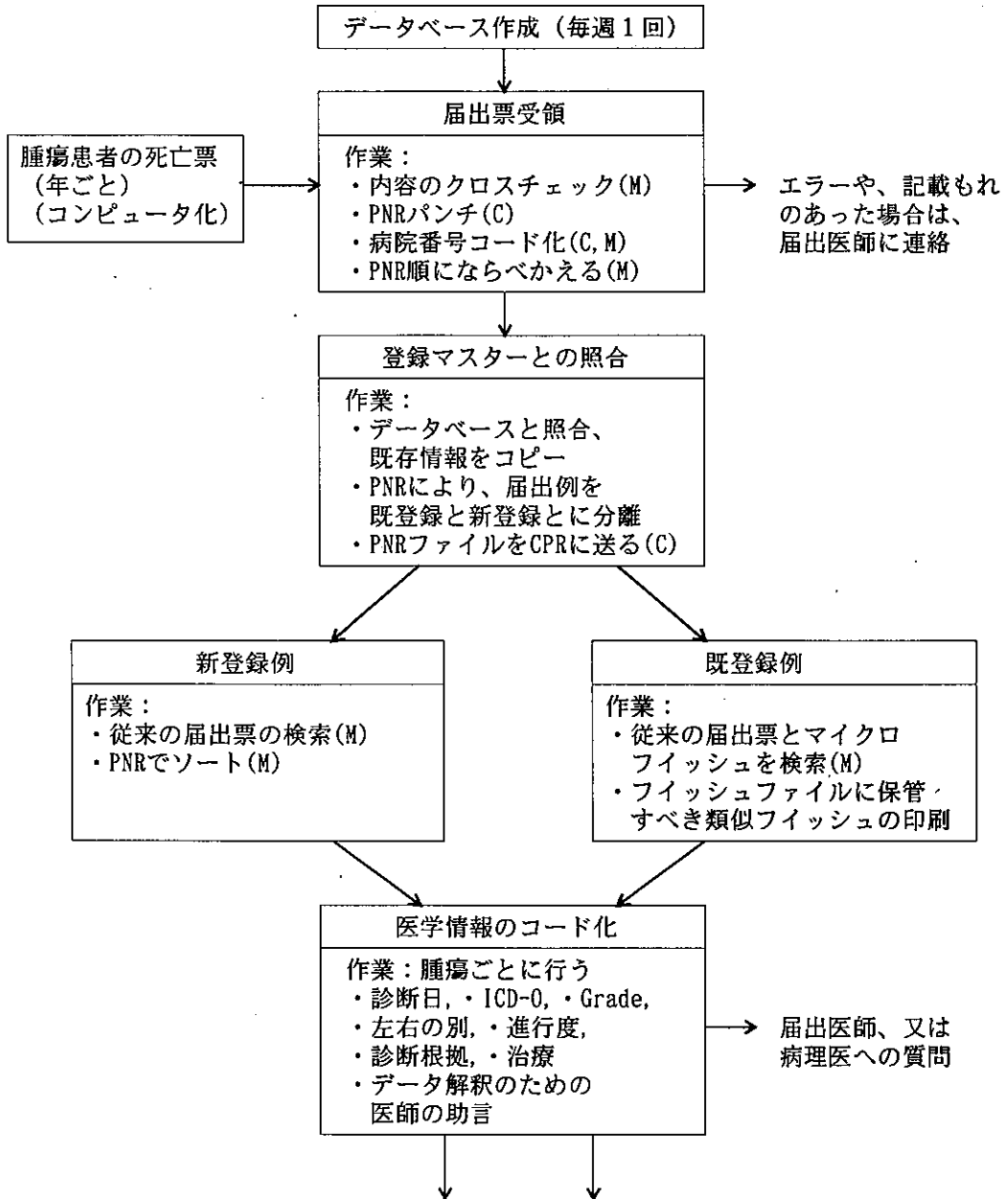
6. THE TREATING HOSPITAL DEPARTMENT

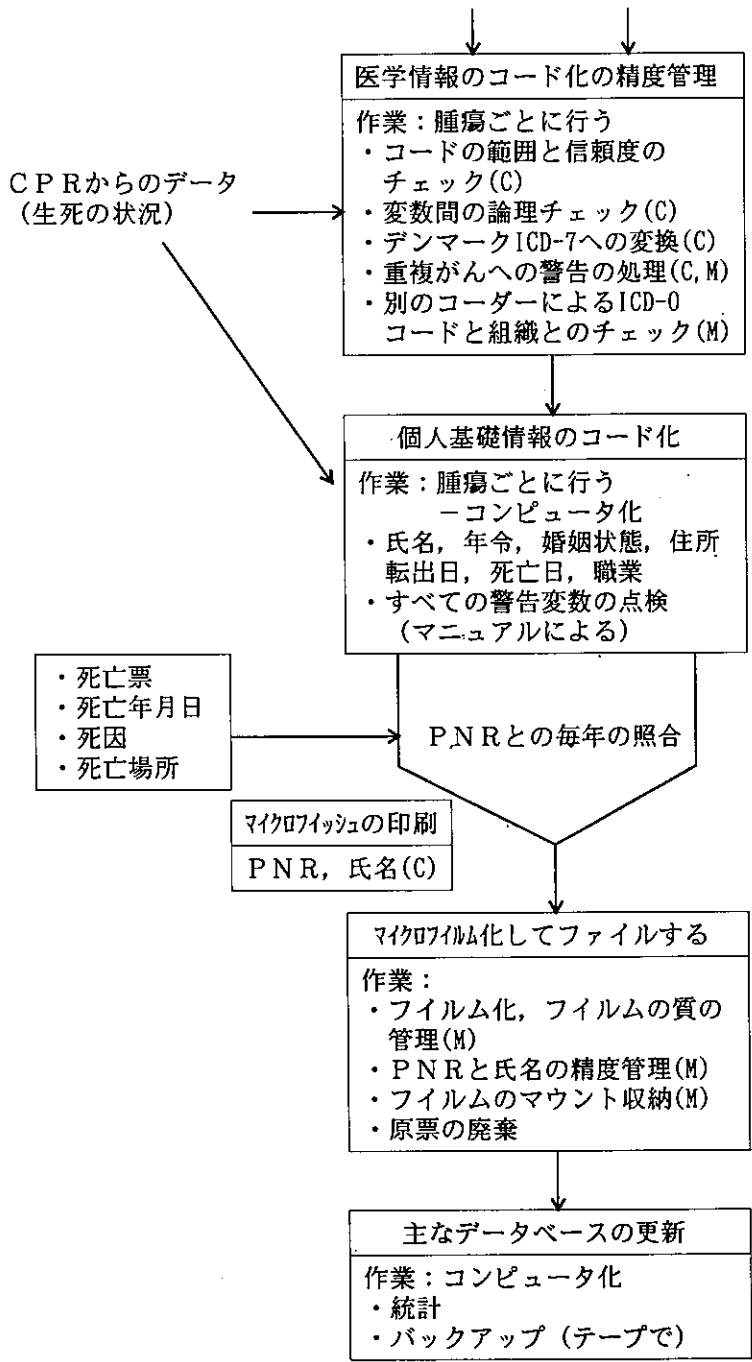
If the tumour was known before autopsy, state the treating hospital department and year of diagnosis.

N.B. If you have any question on how to fill in the form, please contact The Danish Cancer Society, The Cancer Registry, telephone: 01 26 88 66.

図3. デンマークがん登録事業システムの流れ図

注釈：(M) 手作業又は肉眼によるチェック
 (C) コンピュータによるチェック
 CPR 中央人口登録局
 PNR 個人同定番号 (国民背番号)
 MD 医師
 ICD-0 腫瘍学のための国際疾病分類 (WHO, 1976b)





出され、現在の週毎のバッチに移される。PNR による並び変え処理の後、コーダーは、データベースの内容と集められた届出票との間の一致性を慎重にチェックする。CPR（中央人口登録局）とのリンケイジのために PNR（個人同定番号）を有するデータテープが、この段階で作成される。1人のコーダーが、1週間分の届出票のバッチ処理を担当する（届出票受理、コーディング、データ入力、チェック及び修正）。

コーディング

医学情報

特別に訓練されたコーダーが、コーディング・マニュアルの規則と指針に従って、届出票をコード化する。1943年から1978年まで、登録された腫瘍は国際疾病分類第7回改訂（ICD-7）（WHO, 1957）を修正、拡張したものに従い分類された。1978年以降、登録室は ICD-0 分類体系（WHO, 1976b）を用いている。このコードは、診断根拠に関するコードに加えて、膀胱腫瘍に関する Bergkvist の分化度に関するコード（Bergkvist et al., 1965）、一対臓器の側性（左右の別）を示すコードによって拡張されたものであり、従って、ICD-0 コーディングシステムの固有の規則に準拠していない。ICD-0 コードは、コンピュータプログラムによって、修飾 ICD-7 コードに変換される。最後に、診断の後最初の4ヶ月間での病変の拡がりや治療とが記録される。

チェックはオンラインで実行されるものもあるし、一方、1週間分の届出票がコード化された後に、バッチで（一括して）実行されるものもある。コーディングエラーが出現しなくなるまで繰り返し施行される。

この時点で、全ての届出票が、コンピュータ化された情報のリストとともに、部位と形態のようなキー項目の校正のため、別のコーダーに引き渡される。コーディング過程を通して、コーダーは医師の援助を受けるが、この医師は、疑わしいケースを説明するとともに、届出医師や病理医への問い合わせを担当する。私達の経験では、専門家レベルでのコミュニケーションが返事の率と精度を高める。返答の記載が容易に保存できることから、郵便での返事が好まれる。もし電話が使われる場合には、適切な記録を確保するため、返事を書き留めること。

基礎的個人情報

医学情報のコーディングを終えた後、コーダーは基本個人情報をコンピュータに支援されながらコーディングする。患者の公式姓名、婚姻状態、職業、居住地と転出日、あるいは死亡日が CPR から取り込まれる。保存される情報は、腫瘍の診断日現在の情報である。これにより、診断時の年齢を計算することに加えて、例えば患者が診断時に実際デンマーク住民であったかどうかについて、厳密な基準を適用することが出来、また、患者の正しい住所、す

なわち、市町村、郡をとり入れることが出来る。届出票の氏名と CPR 番号が、中央人口登録から取り込んだ情報と一致するかどうか、チェックするために氏名が使われる。これによりコーダーは人数と氏名数の不一致を拾い上げることができる。というのは、正当な根拠に基づく数が、報告施設による誤った名前に由来していることもあるし、あるいは、パンチエラーが偶然 PNR の論理チェックを満足することがあるから。名前は、また、例えば結婚により変わることがあり、コーダーは完全な確認のため CPR で提供される過去の記録を検索し、利用する必要がある。中央人口登録では、職業を重要な項目とは考えておらず、職業はしばしば届出票のみを基礎としてコード化される。自動コーディングの部分には、もしも移送されたデータに不確かさがあれば、コーダーにより確認されなければならないところがある。確認の後、CPR と登録室のデータ間に確認されない不一致がないかどうか、確かめるためのチェックプログラムを走らせる。

精度管理

スタッフの視覚による精度管理

スタッフの視覚による瞬時のかつ継続的な精度管理は、登録過程の重要な部分である。届出票の内容についての視覚検査が、登録作業の初期段階で実行され、脱落部分がないかどうか調べられる。もしも医療情報のコーディングに必要な情報が欠落しておれば、その情報を要請する。届出票の非論理的記載を修正するための情報についても、同様に要請する。腫瘍部位と形態に関するコンピュータ化された情報と実際の届出票情報との一致についての肉眼的チェックは、二重コーディングや二重パンチを行わないので、非常に重要と考える。患者のアイデンティティーと届出腫瘍についての最終の視覚的検査は、マイクロフィルムコピーのマウント及びファイリングと共に行われる。

コンピュータによる精度管理

視覚的精度管理はいくつかのコンピュータによるチェックと警告とにより支援されている。論理のエラーは、コーディング過程を完了するために修正されなければならない。一方、警告は、コード化した情報を変更することなく通過することもある。登録室長のみが、もし必要ならば（例えば女性への性転換手術を受けた男性における精巣がん）非論理情報を入力することができる。値の範囲はパンチ時にチェックされる。チェックプログラムが、割り当てたコードが正当であることを確認し、以下の論理チェックが実行される。

- 診断日は、現在あるいは死亡日と同じかそれ以前でなければならない；
- 性特異的腫瘍は、その性にのみ発生する；
- 一対の臓器は、側性(左右の別)が特定されておらなければならない(不明を含む)；

- 診断根拠が組織診、もしくは細胞診の場合、組織診断が存在しなければならない；
- 診断根拠としての剖検は、患者が死亡している場合のみ受け入れられる；
- 根治手術が実施された場合、手術が診断根拠としてあらわれていなければならない；
- 根治手術は、病変の進展度が限局性または領域性の場合のみ受け入れられる；
- 非論理的な部位と形態学の組み合わせは、検査の後だけ受け入れられる。

最終チェックは警告のプログラムである。つまり、可能であるが、もっともらしくない事象をチェックする。もしも複数の腫瘍が同一の ICD-7コード(3桁レベル)を與えられている場合、二重登録があり得るとい警告が出る。同様の警告が、近接部位の重複がんや ICD-0の部位と形態学の稀な組み合わせの場合にも出る。個人情報のチェックプログラムは、CPRとコードした情報間の不一致を検索する。しかし、もしもその情報を、コーダーが「視覚的検査時に受け入れ」とキーを叩くことによって受け入れておれば、それ以上の警告は出ない。

ファイリングと更新

届出票のバッチが上述の手続きを通過した後、登録室のメインのデータベースが新規情報で更新される。更新結果に関する要約統計が、拒否又は削除された PNR と項目のリスト類と共に、ルーチンに出力される。

マイクロフィッシュシートが、全ての新規患者のキー同定項目、PNR、名前と共に印字される。全ての届出票がマイクロフィルム化され、フォトグラフィックコピーが、1人あたり15のマイクロフィルムが利用できるスペースを有するマイクロフィッシュに保管される。この作業の間に、フィルム、原票とマイクロフィッシュの PNR、及び氏名がクロスチェックされる。もしもこれらが不一致の場合には、手作業でケースに旗を立てられる。もしエラーが見つければ、届出票と変更が、更新処理のために登録システムに「郵送」される。エラーが検出されず、撮影したフィルムの出来上りもよければ、元の届出票は廃棄される。

スタッフ

登録室は、一人のセクション長（医師）によって統括されており、セクション長はスタッフがシステムを運営するのを監督するとともに、既存の登録システムを改善するために指導性を発揮すること、コンピュータセクション、及び国中の臨床家や病理医と協力すること、などに責任を負っている。

1週間に約1,000通の届出票が届く。これらは、5人のコーダー、4人の事務職員とプログラマー1人から成るスタッフによって処理される。基本的なコーディングを実行する事務職スタッフは、この目的のために特別に訓練される。全ての届出ケースのコーディングと分類は、2名の医師によって監督されている。他の専門職スタッフは、CPRより受理した個人情報

報のコンサルタントとして働く。

コンピュータシステム

登録システムは、データベースシステム SIR (1985) を利用して、他の言語による付加的プログラムも追加した上で、構築されている。作表の目的で、登録の SAS (1985) 版が維持されている。変化を受ける事柄ではあるが、データベースの大きさとして、週あたり1,000通の届出票のバッチが選ばれた。メインの登録データベースとのオンラインよりも、むしろ小規模データベースを用いる作業の主要な利点は、ハードウェアやデータベースソフトウェアにおける安全性と、エラーについてのより小さな圧力である。バックアップ作業も、より小規模なデータでは時間が短くてすむ。メインの登録データベースは、処理された毎週のバッチで毎月更新される。コンピュータ化した登録システムは、PRIMEコンピュータ 9955 上を走る。メインの SIR 登録は、200メガバイトのディスク容量を占有している。全体の登録システムは、プログラム、記録、トランザクションを含め、700-800メガバイトまで占有する。

マニュアルと記録

コーディング・マニュアル

コーダー相互間で、また、時代を通じて、コーディングが比較可能な状態に保てるように、詳細なコーディング・マニュアルが作成されている。ICD-0マニュアルは、部位と形態学のコーディングの基礎である。このマニュアルを補足するのが、全ての変数についての項目別マニュアルで、コーディングのための一般的な、また特殊な規約を示している。マニュアルは、登録システムの中で受け入れられるコードを全て維持している。更新の日とこれまでにコード化した情報に対してとる行動について（もしあれば）、明確な記述でマニュアルに述べられていなければ、いかなる変更も受け入れられない。このマニュアルは、特殊なコードに関して機械で読める記録によって裏書きされている。

データ処理マニュアル

コンピュータシステムの種々のステップについて、マニュアルが存在する。マニュアルは、流れ図（図3参照）で概説されたフローの順に従って作られており、データ入力とチェックに関するプログラムを呼び出し、走らせる方法を説明している。マニュアルは機械で読める形で利用できる。

登録システムの記録

詳細な記録と厳密な規約が、常に実行されてきたわけではない。しかし、登録システムに

おける全ての手法と変更についての詳細な記録が、現在手に入る。記録は、プログラム、コード、及びコード間の変換について、詳細に記述している。初期の操作手順に関するかなり正確な記録が、過去に振り返って作成された。発見されたエラーは、全て修正され、記録された。現行の登録システムのエラーを避けるために、全ての変数の値とラベル（すなわち定義）が単一データベースである記録データベース（DOK-DB）に維持される。このデータベースは全登録システムの全プログラムで使われている。コードの値やラベルの変更、修正を実行するときは常に、このデータベースは修正される。変更内容のコンピュータ出力表が、更新日や変数を変更した職員のイニシャルとともに作成され、すべての更新時に視覚的にチェックされる。使われた DOK-DB の改訂版番号は、主登録簿の腫瘍記録に記録される。とりわけ興味深いことに、ICD-0 コーディングシステムとデンマークの ICD-7 システムとの間の変換は、DOK-DB 内にもっているテーブルで行われる。未だ遭遇していないが、変換の変更に対するアクセスと、ICD-0 の部位／形態学の組み合わせに関する新しい変換の追加に対するアクセスは、ともに容易に実行可能である。

データ保護と機密性保持

法律により、また PNR 番号を利用することより、デンマーク登録室は、保健大臣によって予め定められた厳重な規則を遵守している。個人についての全ての作業が検査のために日誌に記録され、保管される。個人に関する情報は、請求者（たち）が登録作業規則に記載されている基準を満たす場合に限って手渡される。医師からの要求や説明がなければ、いかなる人も登録情報から情報を得ることができない。データについての研究は許可されるが、届出医師の同意なしに個人と接触することはできない。

登録からのアウトプット

キーとして PNR もしくは出生年月日、そして第2の選択として氏名を用いて、個人についてのコード化した全情報をディスプレイ上で直接検索することが可能である。SAS コンピュータパッケージを利用して、コード化した変数に関する作表や、その他の変数とのクロス集計表を容易に作成することが出来る。部位別年齢調整罹患率、及び、年齢別罹患率は、コンピュータ化した形で手に入る。ルーチンのデータ、すなわち、年次別、性別、部位別、年齢別、そして地区別の患者数が、年齢調整罹患率（世界標準人口）と共に出版される（デンマーク対がん協会、1987）。さらに登録室は、コード化した情報の妥当性に関する成績、及び各年末時点での有病数を作表する。

結論

登録室は、個々のがん症例について、最小限の情報を要求する。そして、全ての情報は明瞭な記述で提供されなければならない。登録室は、登録室の外で行われるコーディングを頼りにしていない。全てのコーディングと全てのデータ編集作業が、厳密に記述した規則に従い、中央で実行される。このことは経時的なデータの比較可能性を高く保つ上で必要と信じられている。中央化したコーディングは、多くの資料源から情報を収集するという利点を有しており、ケースの分類とコーディングが、全ての情報を考慮して解決できる。中央でコーディングと分類を行うことにより、登録室自身、がん登録過程につき込む専門的技術レベルと努力に責任を負う。登録室は、それ故、疫学研究やがん統計に対する観点をもって、がん登録にとって重要な事項に努力を向けることが出来る。また、臨床的観点から個々の患者を扱う上で重要な分野については、努力を減らすことが可能である。

登録の成功と質にとって最も重要なことは、集積したデータの活用である。この点において、デンマークがん登録はデータの出力を促進するように、また、罹患データが綿密な研究にとって堅固な基礎をなす領域を指し示すことにより、臨床医や研究者がデータを活用するように、勇気づけ、鼓舞するように努めている。

(翻訳：藤田委由、津熊秀明)

付録3 (b) テムズがん登録室

R. G. Skeet

緒言

テムズがん登録室は、カバー人口（1300万以上）が西ヨーロッパで最も大きいがん登録室である。テムズがん登録室は、1958年に南メトロポリタンがん登録室として始まり、1974年に南テムズがん登録室になり、1985年に2カ所の北テムズがん登録室と併合した後、現在の形になった。

収集される情報

テムズがん登録は、基本の細小システムと、膨大な量のデータを複合のデータセットで持つ高度詳細情報システムとの、折衷となることをねらっている。データ項目は、有用なことが明確であり、第一次資料（通常は病院記録）の正確さと完全性がある程度保証され、検索と解析が可能な形でコンピュータに蓄積できること、を基準に選択された。殆んど全ての症例で用いられる登録様式を、図1と図2に示す。

同定項目の細目

新規登録票を既登録索引と照合し、二重登録が避けられるように、患者姓名、性、出生日、住所が記録される。郵便番号は必ずしも記録されるわけではなく、欠損値の処理に特別な方法が使われる（下記参照）。国民保健医療計画番号は、後述するように、生存患者のフォローアップに大変有益であるが、不運なことに病院では記載が乏しい。結婚前の旧姓と出生場所とは、国民保健医療計画番号が不明の場合、がん登録と国民保健医療計画中央登録とを照合するときに、疑問のあるケースの解決に役立つ。

病院の細目

単に意見を求めて紹介され、主たる治療が他施設でなされるのでなければ、患者が受診した各病院の詳細が記録される。病院名、初診日、病院番号、診察したコンサルタントを記録する。

臨床の細目

腫瘍部位および組織型の記録と、診断時の臨床病期の記録とに、あらゆる努力が払われる。TMN 分類、もしくは他の国際病期分類が病院記録にあれば、これを記録する。固形腫瘍につ

外科手術（記入欄は4段あり）			
年月日	病院名	手術名	手術結果 (治癒、非治癒、不明)
放射線治療（記入欄は4段あり）			
開始日	終了日	病院名	使用機器 総線量 分画 1-2-3-4-5
その他の Radiotherapy（記入欄は4段あり）			
年月日	病院名	薬品名	治療回数 1-2-3-4
ホルモン、化学療法（記入欄は各1段のみ）		その他のがん（既往）（記入欄3段あり）	
年月日	病院名	部位	病院名 年 TCR No.
最初の化学療法 最初のホルモン療法			
その他の治療法（記入欄は3段あり）			備考
年月日	病院名	治療内容	
最終生存確認日	患者が死亡している時		
年月日	死亡年月日	剖検（あり、なし）	
	死亡場所		

図2. テムズがん登録室の登録票－裏側

いては、第6章データ項目23で記述した、単純な病期分類体系を利用する。臨床記録が完全でない場合でも、通常はこれらのデータを抽出することが可能である。外科手術を受けた症例については、確定情報という観点で、データ項目が繰り返すとられる。

治療の細目

地域がん登録における治療細目の記録については、議論のあるところであり、実施される場合は常に妥協の産物である。テムズがん登録室のみが、初期治療計画で実施された治療の記録を試みており、これには、診断後3カ月以上経って始めた治療は、予め計画されていた場合を除き、通常、含まれない。外科手術については、手術の性格と手術の完全性についての何らかの評価が記録される。外部照射療法（この詳細は病院記録に殆んど常によく記録されている）については、使った装置、腫瘍照射量、分画数、照射部位が記録される。その他の放射線治療については、用いたアイソトープ、照射回数、照射部位が記録される。詳細を記録せずに意味のある方法で、ホルモン療法と化学療法のデータを収集することは、極めて困難であり、それ故解析を困難にしている。実際、テムズがん登録室は、これら治療の事実

のみを記録する。長期間の情報システムにとって安定が重要で、ソース記録に対する頻回の変更は、それが必要不可欠のものでなければ極力避けられるので、将来何か新しい治療法が採用された時にフラッグを立てるように、他の治療法に関する1セクションをもっている。

別の悪性腫瘍

病院記録によって、ある患者が2個以上の原発性がんを発症したことが明らかになった場合、コンピュータシステムの中で、この情報のリンケイジが出来るように、先発もしくは同時発生腫瘍の詳細情報が提供される。

人口動態資料

ケースが抜粋される時点での患者の状態に関する基本的細目が記録される。これらは後に死亡診断書の細目によって補足されることもある。

データ収集方法

登録症例の殆んど全ては、登録室によって雇用されたフィールドスタッフ（研究事務員として知られている）によって、巡回して抜粋される。各員は、基地とする病院をもっているが、近郊の他のいくつかの病院をも訪問し、可能な場合はどこの病院でも（通常は患者を相互に紹介する病院グループ内で）働く。ロンドンにおいては、これら事務員の多くはフルタイムで雇用されているが、首都以外では、遠隔の病院間で不必要な旅行を避けるため、パートタイムで雇用されていることが多い。これら事務員は、登録室のメンバーである故、登録室によって選ばれ、訓練され、監督されている。このことにより一貫性と専門性を出来るだけ高く維持することが確実に出来る。訓練は継続的なプロセスであり、これを促進するために研修日が設けられている。研究事務員は、4つの地域チームに編成され、各員は、それぞれ監督者をもっている。少数の病院においては、自前で研究事務員に相当する職員を備えており、これら職員も登録室の研修プログラムに参加するように勧められている。これらの職員の多くは、がん登録とは直接関連のない他の事務的業務をもっており、がん患者について自らの情報システムを維持することに関心を抱く病院でみかけられる。

毎月、研究事務員は完成した登録様式を登録室に送付し、登録の必要なケースのコンピュータリストを受け取る。これは、例えば、未登録のがん患者について死亡診断書によって明らかになったケースである。彼らは、解決の必要な、例えば、診断が疑わしい症例、詳細が欠落しているか情報が矛盾している症例、等の問い合わせのリストをも受けとる。

巡回フィールドワークのシステムは、託すべきものを多く抱えている一方で、大規模の登録室にとって相当な費用と管理とが要求される。日々の仕事にあたって研究事務員は、監督されない状態で、そしてせいぜい訪問者、最悪の場合は邪魔者としてみられる環境で活動す

る。病院スタッフとの関係は、忍耐強く深められなければならないが、訪問（研究事務）員が病院自身のスタッフよりも、より永続性のある場合には、困難を伴うかもしれない。堅固さは、気転と結びついていなければならない。というのは、ひ弱な研究事務員も横柄な研究事務員も成功しないからである。そうしたスタッフの任用に当たって、パーソナリティは資格と同じくらい重要である。

その他のデータ源

死亡票

多くの他の登録と同様に、テムズがん登録室は、その地域内で死亡した患者、及び、悪性疾患が原死因もしくは寄与死因の者の死亡診断書のコピーを受けとっている。死亡診断書の処理は、図3に概説されている。診断書が既登録患者についてのものであれば、コンピュータ記録が、そのように更新される。病院で死亡した未登録患者の場合には、病院にいる研究事務員に通知され、登録するように要請する。患者が自宅で死亡した場合には、登録室は死亡診断書を記入した医師に手紙を出し、患者が診療を受けた病院の細目を要求する。もしもその患者が一度も病院を受診していなければ、そのケースは死亡診断書のみに基づいて登録される。

組織型報告

コンピュータ化した病理システムの迅速な発達に伴って、テムズがん登録室は、病理コンピュータと登録コンピュータ間とのデータ転送による悪性疾患の予備的届出システムの稼働に向けて準備を進めている。

データのコンピュータ化

登録用紙に記録された情報は、全てコンピュータに入力される。テムズがん登録室は、登録の全データベースがオンラインで維持される精巧なシステムを用いている。すなわち、ディスクに保管され、どのケースでもその詳細を、直ちに端末に呼び出すことが可能である。記録はどのケースでも、番号、名前、もしくは、解析プログラムを利用して、診断名により呼び出せるように、索引化されている。

レコードリンケイジ

入ってくる各々の記録について、最初に行われる作業は、それが既登録患者に関連した情報かどうかをチェックすることである。これは、患者の氏名、性、出生日をコンピュータに入力することで行われ、コンピュータはそれらにより一連の検索を実行する。すなわち、

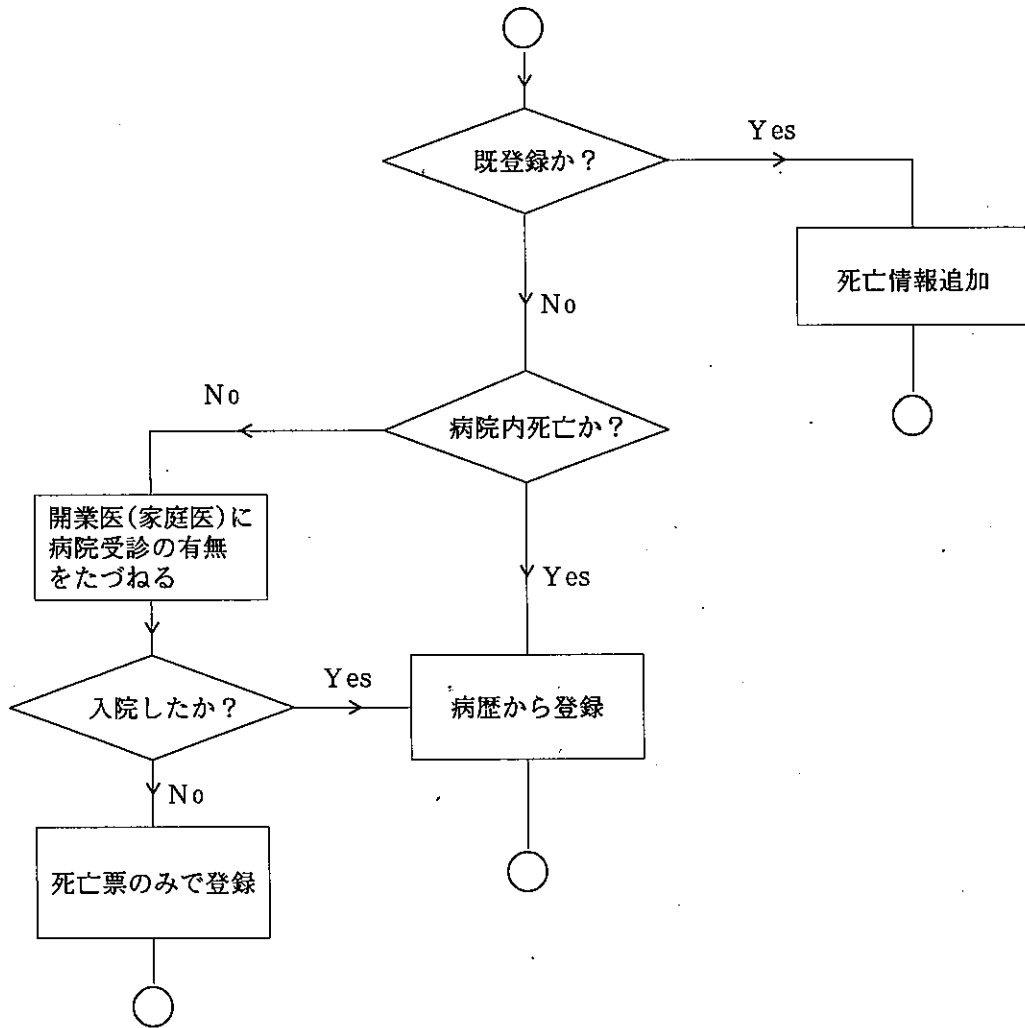


図3. テムズがん登録室における死亡票の処理

最初に直接照合を実行し、次に出生日ウインドウ、つまり一定の誕生日の周辺の類似データを検索し、最後に氏名がこれまでに違って綴られたことがあるケースについては、音声検索を利用して照合する。もしオペレーターがこの患者は登録簿にないと確信すれば、コンピュータはこの患者に、新規の患者番号を自動的に発行する。コンピュータは、登録票の残りのデータを入力することにより、いま登録した腫瘍に、同一番号（腫瘍番号）を割り当てる。その患者が既に登録されていることが分かった場合には、腫瘍（または複数の腫瘍が登録されている場合は複数の腫瘍）の詳細がスクリーンに表示される。この時点でオペレーターは、登録票上のデータが既に登録された腫瘍と関連しているのか、もう1つ別の未登録の腫瘍と

関連しているのか、決定しなければならない。前者の場合、既にある情報に更に詳細な情報を追加するために登録票が使われるが、後者の場合には、新しい腫瘍番号が割り当てられ、そのケースが通常の方法で登録される。このように、患者は唯一の患者番号を持つが、2個以上の腫瘍番号を持つことがある。各患者にとって登録された最初の腫瘍は、患者番号と同一である。このように1患者における個々の原発腫瘍は別々に数えられるが、同じ患者番号を持つことによりリンクされる。

データのコード化

テムズがん登録室ではデータをコンピュータ上、コード化して保持しており、それ故、データは解析のために適切に組織化されているが、全てのデータは用語形式で入力される。コンピュータは入力した用語を自動的にコードに変換し、また、データが呼び出されるときは用語に再変換される。再変換された用語は、当初に入力した用語と完全には一致しないことがある。例えば、オペレータが 'Bile duct carcinoma' を入力した場合、コンピュータは 'M81603' とコードするが、再変換時には 'Cholangiocarcinoma' として呼び出されることになる。この「好ましい用語・類義語」システムは、ローカルな用語や省略形に適合させることが可能な、非常に可塑性の高い辞書を作り上げた。難解であいまいな用語に遭遇した場合に、利用者は辞書を参照することが可能で、必要に応じて新しい同義語が追加される。この方法でコード化されるデータ項目の中には、出生地、病院、コンサルタント、職業、部位、組織、手術、放射線治療装置と放射性アイソトープ、などがある。

整合性のチェック

届出票を格納する前に、コンピュータは一連の整合性チェックを実行し、オペレータは見つかったエラーを修正するように要請される。

郵便番号及び詳細地域のコード化

大英帝国の住所の郵便番号から、住民の居住地区の市町村役場、選挙区、保健区域を、コンピュータ化した表を参照することにより、決定することが出来、これにより、これまで登録室で行われてきた最も難しい手作業によるコード化の幾分かを解消することが出来た。郵便番号が登録票に記録されていない場合には、住所地郵便番号のコンピュータコーディングを専門に扱う部局に、患者住所が磁気テープで送られる。この作業は年4回実施され、再び磁気テープを用いて郵便番号がデータベースにフィードバックされる。

二重登録の後向きチェック

1年間に12万件を越える記録が入ってくる登録室において、ある程度の二重登録が発生することは、殆んど不可避である。これは、オペレータのエラー、同一患者による別の名前の

使用、あるいはその他の要因によって引き起こされる。コンピュータは、同一の郵便番号と、類似あるいは同じ部位コードを持つ2年間のケースリストを、定期的に出力する。1郵便番号は、平均15の居住地住所のみカバーするので、2年間の内に同一のがんが2症例存在する確率は非常に稀であり、二重登録を発見するために慎重な手作業によるチェックが行われる。

問い合わせ状の作成

登録がん症例に関する問い合わせ状の作成は、非常に日常的な処理手段であり、容易にコンピュータ化され、これにより大量の事務的努力が省力化される。例えば、図3で示すように、病院や開業医への問い合わせ状の作成は、全てコンピュータにより実行され、オペレータは単に死亡診断書の詳細をキー入力するだけである。システムは、さらなる情報の必要性を検出し、そして最初の問い合わせ状を印字した後、適当な行動が実行されるまで、適当な間隔で催促状を発生する。

全国がん登録室へのデータ送付

4つのテムズ地域に居住する各症例についてのデータの一部が、政府統計局 (Population Censuses and Surveys、OPCS) によって運営されている全国がん登録室に送付される。データは年4回磁気テープで送付され、England and Wales の他の登録室から送付されるデータと併合される。

フォローアップ

テムズがん登録室は受動予後調査方式を採用しており、死亡したか転出したかが不明の症例は全て生存とみなされる。がんによる死亡は、死亡診断書で通知される (図3参照) が、がん以外の死亡の場合や登録地域の外で死亡が発生した場合は、登録患者の死亡を登録室に通知するのに、別の方法が用いられる。

後者の方法は、OPCS に依存している。この OPCS は、英国がん登録室 (National Cancer Registry) を運営するとともに、大英帝国の全人口の手作業索引を維持している英国保健医療計画 中央登録 (NHSCR) をも運営している。全ての登録患者は、全国登録室で、常時、全国の死亡票との照合により観察されており、死亡がどこで発生しようと、死因が何であろうと、患者の死亡が判明したとき、その旨、関連のがん登録室は通知される。

データ解析と情報の取り出し

テムズがん登録室は、最初にデータを作成するのに用いた辞書とテーブルとを中心にデザ

インされたコンピュータプログラムを、データの取り出しに使う。この出力と入力の機能的関係は、効果的で合理的な情報サービスを維持する上で重要である。殆んど全ての特別な問い合わせは、利用者が使い易いプログラムを活用することにより処理される。その際プログラムの作成は殆んど必要ない。

登録室の組織

登録室は、データ収集（巡回研究事務員）、データ処理（全ての入力作業を行う VDU オペレータ）、及び、情報と研究（データ出力操作を実行する科学スタッフ）の、3つの機能単位に組織化されている。これらの単位が、互いに独立しているとするのは誤りであり、これらの間の、よいコミュニケーションが必要不可欠である。これは、非公式な日常を基礎として実行されるが、これらの単位の責任者と4人のデータ収集監督者との公式会議が毎月開催され、進捗状況をレビューし、現実の、あるいは将来起こりうる問題点を議論する。組織内の効果的なコミュニケーションのつながりは、効果的で高精度の情報システムを維持する上で、極めて重要である。

（翻訳：藤田委由、津熊秀明）

付録3 (C) オンタリオ州におけるがん登録

E. A. Clarke, L. D. Marrett and N. Kreiger

背景

オンタリオがん登録室 (Ontario Cancer Registry; OCR) は、オンタリオ州全体をカバーする地域がん登録室である。オンタリオ州はカナダで人口密度の最も高い州であり、1986年の人口は910万人 (Statistics Canada, 1987)、面積は百万平方キロメートルを越える。人口の82%は都市部に居住しており、都市部の大部分は、州の南部に位置している。住民の80%はカナダ生まれであるが、多様な人種性を示し、大部分がイギリス系、フランス系、イタリア系、ドイツ系である。

OCRはオンタリオがん治療・研究財団 (Ontario Cancer Treatment and Research Foundation) によって運営されている。オンタリオがん治療・研究財団は、1943年に、州における「がんの診断、治療、研究計画の設立」を目指したがん制圧法 (Cancer Act) の制定に基づいて組織された。この法令は、「放射線治療、さらに外科治療以外の効果の高いがん治療を中央に集中させる」という州の検討委員会の提案に従ったものである。従って、外来患者を対象とした放射線治療を提供するために、州の主要な市に地域がんセンター (Regional Cancer Centres; RCCs) が設立された。さらに、マーガレット王女病院 (Princess Margaret Hospital; PMH) を統合して、オンタリオがん研究所が1958年にトロント市に設立された。RCCsとPMHとを合わせると、州のがん患者の放射線治療の全部と、がん患者の化学療法及び相談サービスの50%とが提供されている。

OCRを含めたオンタリオがん治療・研究財団は、基本的にオンタリオ厚生省 (Ontario Ministry of Health; MOH) に支援されている。患者の治療は公的に給付されている。オンタリオでは、約95%の住民が、政府の健康保険計画に含まれている。州の外で医療を受ける住民もいるが、オンタリオ州外からくる入院患者治療の請求率は1%未満である。そのような請求の大多数は、州の境界近くに住んでいる住民によってなされている。

1943年のがん制圧法は、「がん症例の十分な届出と、データの記録と編集」計画も含んでいた。がんは、今も、オンタリオでは法律に基づいて報告される疾患ではないが、1943年からのがん制圧法の改正により、オンタリオがん治療・研究財団にがん症例を報告するヘルスケアシステムに所属する組織及び個人は、法的に保護されている。これらの改正によって、OCRの情報は、疫学的、医学的研究に利用できる。さらに、州の各病院は、全ての退院患者に関する診断情報を、MOHに転送することを要求される。MOHは、これらの情報を行政目的に使い、がん患者のデータのコピーをOCRに提供する。従って、入院患者に関しては、ある程度、報告の義務が課せられている。

がん登録の手順

OCRには、1964年以降に診断されたがん患者が含まれているが、1972年に登録方法が大幅に改正された。この報告では、1972年以降の方法について述べる。以前の方法の詳細に関しては、オンタリオがん罹患データの初期20年に関するモノグラフに掲載されている(Clarke et al., 1987)。OCRでは、黒色腫以外の皮膚がんは、登録していない。

OCRは、種々の情報源から提供される、元来はがん登録以外の目的で作成された記録から、がん登録ファイルを創生する。コンピュータによる記録連結システムが、これらの情報をまとめ、同一個人に属する多数の記録を連結する。その後、症例決定システム (the Case Resolution system) というコンピュータ化された規則が、連結された記録に適用され、各々のがん症例に、適当な部位、組織、診断日と診断方法、居住地、およびその他の情報を割り当てる。これらの方法は、オンタリオがん治療・研究財団の2つの部門、即ち、疫学と統計情報システムとの協力の結果である。

データ源

OCRのファイルを作成するために、4つの主要な情報源が採択されている。

がんと診断された退院記録

がんの記載のある病理報告

がんが原死因の死亡診断書

RCCsおよびPMHに紹介された患者に関する記録

(1) 病院退院記録

がんの記載のある入院患者の退院記録が、MOHによってOCRに送られる。これらは1975年までは文書で提出されていたが、それ以降は磁気テープで提供されている。1978年に、MOHは、各病院に対し、全ての退院抄録を、独立機関である病院医療記録研究所 (Hospital Medical Records Institute; HMRI) に提出するよう、要請した。HMRIの抄録書式は、最高16個までの退院診断を記載することができる (1978年までの退院記録では1個の診断しか記載できなかった) が、これらの抄録には姓あるいは名が含まれていない。(ある種の編集も含めて) データ処理加工の後に、HMRIは結果ファイルをMOHに転送し、MOHで氏名とオンタリオ健康保険番号 (Ontario Health Insurance Plan; OHIP) が付加される。この統合ファイルから、退院診断のひとつにがんのある記録から成るファイルが作成されて、毎年OCRに転送される。現在、毎年約10万件の退院記録が受け取られている。

(2) 病理報告

1973年に、州内の全病理検査機関は、がんの記載のある記録のコピーを提出することを要求された。1980年から全てが応じている。OCRが受け取る毎年の病理記録の数は、1973年の

15000件未満から近年の約5万件に、大幅に上昇している。参加している病理検査機関からOCRに紙の記録が提供され、OCRの職員がコード化している。

(3) 死亡

OCRは、オンタリオ住民の全死亡に関するデータを機械で読める形で持っている。1972-80年は、これらのデータは Office of the Registrar General of Ontario との特別な調停のもとに、Statistics Canadaから提供された。1981年からは、Office of the Registrar General of Ontario が直接コンピュータ・テープをOCRに毎年提供している。原死因は、Office of the Registrar General の訓練された疾病分類士によってコード化される。がんが原死因と考えられる全死亡がOCRのファイルにとり込まれる。オンタリオ住民のがん死亡数は、1972年には11500例、1986年には17000例であった。

(4) 治療センター

まず、がん患者の最少限の情報を記録した抄録カードが、各RCCおよびPMHで作成される。RCCsからの抄録カードは、さらにデータを抜粋し、コード化するために、OCRに転送される。1972年から1981年の間に、これらのカードはPMHおよびRCCsで徐々に作成されなくなり、かわりに機械で読める形式で、適切なデータがOCRに転送されるようになった。RCCsでは、その地域で診断されながら、センターに紹介されなかった症例に関しても、RCCで維持しているがん登録のために抄録カードを作成している。これらのカードは、1976年にRCCsのがん登録が中止されるまで、採録とコーディングのためにOCRに転送されていた。OCRでは、RCCsとPMHから毎年約2万件の記録を受け取る。

コーディング、データ入力、データ作成

初期(1972-1975)にOCRに提出された記録は、がんが原死因として記録されている死亡診断書を除いて、全部OCRで集中的にコード化され、コンピュータに入力されていた。1975-1977年の間に、病院退院記録はMOHでコード化されるようになった。1978年からは、オンタリオの病院の病歴室でコードが付けられている。これらのデータは1975年から磁気テープでOCRに送られている。受動的な登録方法を採用している限り、多くの場合、コーディングに関する質の制御のための公式な方法を設立するのは不可能である。

病理記録は、常にOCRの職員の手でコード化し、入力してきた。これらは、OCRで以前にコード化されていた他の記録と同様に、定期的な質の評価の対象となる。難しい症例はコーディングスタッフの間で回覧し、医療スタッフとの定期的な会合で討議される。

RCCsおよびPMHからのデータは、その設立時から同一の形式で集められてきた。これらのセンターにおける記録のコンピュータ化に伴い、コード化は、それぞれの病歴職員によって行われるようになった。各RCCとPMHの病歴室長が、年に2回、コーディングと他の質の制御に関する問題について討議するために、会合をもっている。RCCsは、病理記録とがんの臨床的

記録とをOCRに送っている。これらの記録は再コード化され、違いがあった場合は必ず、RCCとOCRとで討議の後に修正される。

データ入力時の定期的な質の制御は、OCRの全データについて行われている。オンラインで入力された記録からサンプルを抽出して、定期的に再コード化を行い、再入力することで点検が行われている。データ入力システムは、一定の変数（例えば、患者の姓、診断日、疾患の部位）が必ず入力されていることを要求する。データが入力されると、妥当性、一致性、もっともらしさが検査される。磁気テープで受け取ったデータも、同様に点検の対象となる。しかし、これらはバッチプログラムで実行される。妥当性の検査では、本質的に間違えているデータ（例えば、13月、32日）を拒絶する。一致性の検査では、2つ以上のデータ項目を比較し、矛盾するものを記録する（例えば、男性の卵巣がん、誕生日以前の治療日）。もっともらしさの検査では、起こりそうもないが、全くないこともない状況で、エラーの可能性が高いものを記録する（例えば、110歳の患者、5歳の前立腺がん患者）。これらのもっともらしさの検査では、手作業でチェックし、必要があれば修正する。コード化されたデータ（例えば、居住地、病院、出生地）は、OCRが確認の目的で特別に作成した表と比較する。最後に、数値データは、チェックディジットで検査する。

全記録の部位は、1979年までは国際疾病分類第8版（ICD-8）（WHO, 1967）でコード化されていたが、1979年からは第9版（ICD-9）（WHO, 1977）を用いている。前述したコンピュータによる編集に加え、全てのICDコードはデータ編集過程でICD-9に変換される。1979年以前は、組織はMOTNAC(Percy et al., 1968)でコード化されていたが、1979年からは国際疾病分類腫瘍学(ICD-O)（WHO, 1976b）を用いている。MOTNACコードも、同様にコンピュータでICD-Oの形態学(M)コードに変換される。

データ連結

いったん、元ファイルが加工されると、一人の個人に帰属する全ての記録は、連続したコンピュータ連結システムによって、一つにまとめられる。この大容量のデータを一つにまとめるために、OCRでは、Statistics Canadaがカナダの国立がんセンターの疫学部門と共同開発したGeneralized Iterative Record Linkage System (GIRLS)に基づいた、優れたデータ連結システムを開発した(Howa & Lindsay, 1981)。

オンタリオでは、一生を通じて個人を識別する、保健あるいは政治システム上の個人番号を持っていないので、連結は、氏名、誕生日、OHIP番号、診断した病院と病院のカルテ番号等、のいくつかの同定項目に基づいて行っている。OHIPは世帯に割り当てられていて、その世帯内の個人を識別しない。子供が18歳になる（もし大学に就いているならば21歳）と、新しいOHIP番号が割り当てられる。転職、離婚に伴って、個人に割り当てられているOHIPは変更される。

現行のコンピュータ連結は、数段階から成っている。まず、New York State Identification Intelligence System (NYSIIS)コードが作成される。NYSIISコードは姓の音声編である。同じNYSIISコードを持つ記録だけが、連結の可能性のある記録として比較される。従って、似たようなスペルでも、NYSIISコードが異なる姓は連結の機会がない。同じNYSIISコードを持つ記録がひとつのポケットを構成し、そのポケット内で考える全ての組み合わせに対して比較を行う。2つの記録を比較する際に、数字のスコアあるいは重みが各項目に割り当てられる。比較された項目の重みの合計が大きくなればなるほど、2つの記録がシステム上で同一個人として連結される可能性が高くなる。頭字語（各語の最初の英字をならべて作った語。例えば United Nations はUN）である GIRLS の I は、iterative（反復する）から由来しており、この重みの割当過程が一回以上繰り返されることを示している。システムでは、より正確な重みを割り当てるために、以前の観察結果を用いている。

各々の連結（即ち、まとめられた各記録のペア）は、重み合計の大きさに基づいて、3つのカテゴリー、即ち、確定、疑い、棄却、のうちのひとつに分類される。連結ファイルにおける重み合計の分布は、通常2峰性であり、高いところ（確定、即ち真実の連結の可能性が高い）と、低いところ（棄却、即ち真実の連結の可能性が低い）の周りに群を作っている。重みの中央部分は、連結の疑い、即ち、そのペアが同一個人に関係していることが不確かであるものを含む。この範囲（灰色部分）の連結記録については、連結に用いられなかった追加データも用いて、OCRの保健記録職員が検討する。一例をあげると、術前に生検（バイオプシ）が行われたという職員の推測が完全な病理記録に含まれていた情報により確認された。灰色部分の採用もしくは棄却に関する決定が行われ、結果が連結ファイルに入力される。これらの手作業による決定は、間違った連結を採択したり、真実の連結を棄却する数を減少させるが、これら2種のエラーはまだ発生する。灰色部分の大きさは、連結されるファイルに応じて異なるが、連結可能性のある組み合わせの2-12%が手作業で決定されている。

元ファイルの連結は、図1に示した順に行われる。毎年、同一患者の複数の入院をまとめるために、退院記録内の連結が行われる。それから、病理記録がこれらのまとめられた記録に連結される。なぜなら、ほとんどの病理記録は、ひとつの入院に関係しているからである。この退院-病理連結ファイルは、新規患者であること（有病者でないこと）を同定するために、次いで、前年までの罹患と連結され、仮の罹患データを作成する。2年あるいは3年毎に、がん死亡と、RCCsおよびPMHからの記録とが、これらの仮の罹患データと連結される。RCCおよびPMHからの記録は、データの質、とくに部位と組織の質を改善するが、これらの最終連結により追加される症例は殆んどない。

最終的に、NYSIISコードによって作成されたポケット以外のポケットを用いて、全ファイル内の連結が行われる。この連結によって、異なったNYSIISコードを持つグループが比較されて、同じ個人に帰属する可能性のある記録が連結の機会を持つ。この連結は、3つの異なる段階から成る。各段階で作成されるポケット内では、他の連結と同様に、可能性のある全

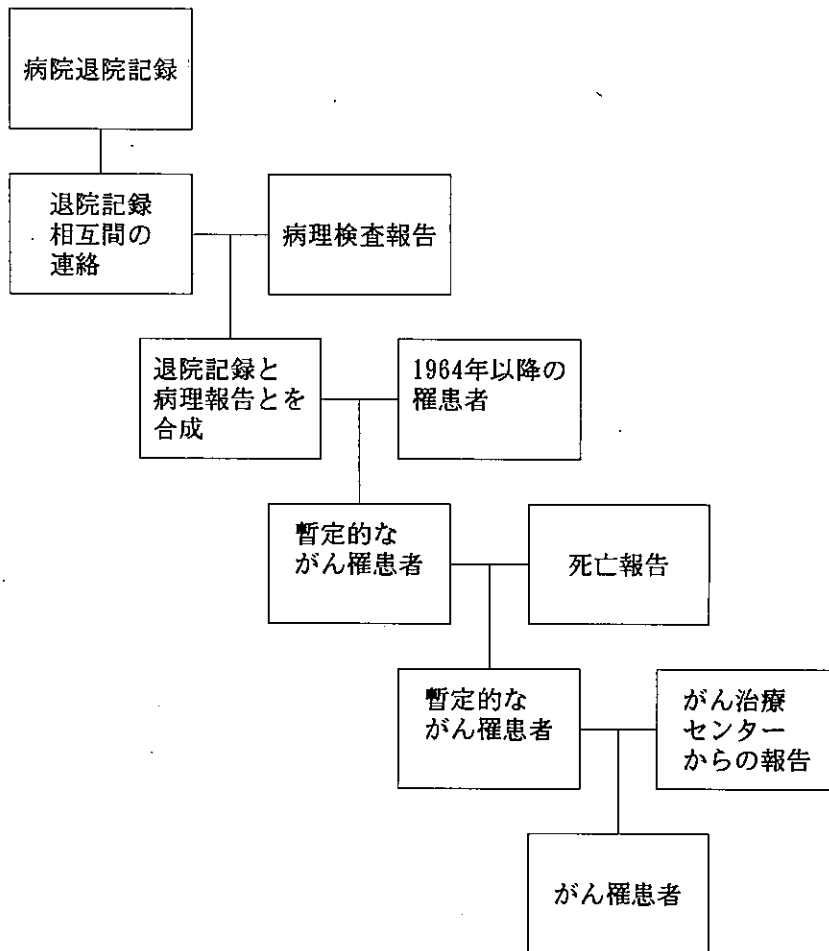


図1. 情報源としたファイルの連結の順序

てのペアの比較が行われ、重みが割り当てられる。第1段階では、ポケットはOHIP番号と性別とを用いて行われる。第2段階では、ポケットは、生年と名の最初の3文字で作成される。第3段階では、姓の最初の3文字を利用する。第1段階で重みの高い連結は、その後の連結に含まれない。この3段階の連結で生じた灰色部分は、通常の連結と同様の方法で判定する。これらの連結により、姓名のスペルミス、転記ミスを減少する効果がある。

ほとんど全例の患者が、複数の情報源に由来する記録を持っている。1人の患者に関する全ての元データの中から、患者を同定するのに最もよい情報（例えば、姓、名、生年月日、性別）を複合させた患者同定記録を作るために、OCR職員が、一連のコンピュータプログラムを開発した。これは、以後の連結ステップで使用される。これらのプログラムは、以前では発見されなかった間違っただけの連結の結果、生じたかもしれない各々の元データ間の矛盾をも発

見する。これらの矛盾は記録され、OCRのコード職員によって吟味され、指示に従って修正される。

部位、組織、およびその他の情報の割当

個々の患者に連結された元データの群は、2つめの大きなシステム、即ち症例決定システム、によって加工される。これは、OCR職員によって開発されたコンピュータモジュールから成り、患者の元データに医学的論理を適用して、適切な部位、組織、診断日を決定する。なぜなら、これらは記録の間で異なるかもしれないからである。

症例決定システムでは、ICD分類あるいは分類群の最も特異的な部位を決定するために、1人の患者に帰属する全ての記録におけるがんの部位を吟味する。最も特異的な部位コードだけが生き残って、次の過程に進む。従って、もしある記録で、「消化器がん、部位不明(ICD-9 159.9)」とあり、他の記録で「胃がん、部位不明(ICD-9 151.9)」、また他で「幽門の悪性新生物(ICD-9 151.1)」、他で「幽門前庭の悪性新生物(ICD-9 151.2)」とあれば、151.1と151.2が最も特異的であるので、その2者だけを残す。

もしこの過程で、たったひとつのコードだけが残れば、それが原発部位であると見なされる。2つ以上が残り、かつ、ICDの4桁目だけが異なる場合(例えば、151.1と151.2)は、最も信頼性の高い記録の部位を採択する。この目的のためには、RCCおよびPMHの記録が最も信頼性が高く、次いで病理記録、それから退院記録、最後が死亡診断書となる。

もし2つ以上の3桁コードが残ったら、その患者の各々の記録の組織コードを比較する。Berg(1982)が作成した分類の改訂版(表1, OCRで改変)では、同じ組織と考えるべき組織がグループごとにまとめられている。組織型の項目のデータがないか、あるいは組織所見が「新生物、詳細不詳」(ICD-O M-800)か「顕微鏡的確定なき腫瘍」(ICD-O M-999)の記録は、全ての組織群に含まれる。さらに、組織型が「がん、詳細不詳」(ICD-O M-801)と「未分化がん、詳細不詳」(ICD-O M-802)の記録は、肉腫、リンパ腫、白血病を除いて、全ての組織群に含まれる。この表に載っていないICD-O Mコード、例えば、「粘表皮腫瘍」(ICD-O M-843)は、他のどれからも異なった組織をもっていると考える。

OCRでは、明確に再発ではないと表示されていない限り、同一個人における2つめの部位のがんは転移と考える。従って、第2がんを登録する規則は保守的である。同一個人で、2つの原発部位が登録されるためには、部位がICDの3桁レベルで異なり、さらに、2つの部位の組織型が、表1の中で異なるグループに属していなければならない。この規則の唯一の例外が、乳がんである。対側に乳がんが転移することはほとんどない。もし1つめのがんと対側の乳がんとが元データでみられたら、たとえ組織型が連結されたデータの他の原発部位と同じグループであっても、原発部位として登録する。他の原発部位から決定された症例も登録する。

表1. オンタリオがん登録室が、部位の割当において同じ組織型と考える組織型分類

用語 ICD-0 M	番号 ICD-0 M ^a
扁平上皮がん	807-808
移行上皮がん	812-813
腺がん	814, 816, 818-823, 825-838, 857
付属器がん	839-842
嚢胞性、粘液性、および漿液性がん	844-848
導管性がん	850-854
生殖器特殊がん	859, 860, 862-867
旁神経節およびグロムスがん	868-871
黒色腫	872-874 876-878
肉腫および他の軟部組織のがん	880, 881, 883-886, 889-892, 899
奇形性がん	908, 909
血管およびリンパ管のがん	912-915, 917
骨肉腫、軟骨肉腫および歯原性がん	918, 919, 922-927, 929-933
その他の腫瘍(松果体腫、脊索腫及び顆粒細胞筋芽腫)	936, 937, 958
神経膠腫	938-948
神経上皮腫性新生物	949, 950
神経鞘性腫瘍	954, 956
リンパ腫およびホジキン病	959-966, 969-972, 975
白血病	980-994

a : 国際疾病分類 (ICD-0) (WHO, 1976b) の形態コード

もしICD3桁レベルで異なる部位が、表1の同じグループの組織型を持っているならば、転移とみなし、最も信頼性の高い記録(前述と同様)に報告されている部位を原発部位と考える。データ源の信頼性が等しい場合は、最も信頼性の高い情報源からの、より特異的な部位コードを全て含む広いコードを、原発部位のコードとして採択する。例えば、横行結腸の腺がん(ICD-9 153.1)と直腸の腺がん(ICD-9 154.1)が、同等に信頼性高い情報源から報告されているならば、「大腸、部位不明」(ICD-9 153.9)を割り当てる。

OCRでは黒色腫以外の皮膚がん(ICD-9 173)は登録していないが、もし適切ならば、信頼性の低い情報源の部位と組織とを無効にするために、この部位の記録を用いることがある。従って、もし退院記録ががんの部位として「口唇の悪性新生物」(ICD-9 140.9)の診断を示しているも、口唇の皮膚(ICD-9 173.0)の病理記録があれば、登録しない症例となる。

特異的な割当ができなければ、原発部位は「部位の明示されない悪性新生物」(ICD-9 199)に決定される。これは3つの方法で起りうる。1つめは、記録されている唯一のコードが199である場合。2つめは、信頼性が同等な情報源で、組織型が同じで、異なる器官に原発部位があった場合。3つめは、原発部位の情報がなく、2つ以上の続発部位(ICD-9 196-198)が特定されていた場合。

もし、2つ以上の原発部位が同定されると、各々の元データを吟味し、この診断に最も適当なものに関連づけ、各部位別に全てのデータをまとめる。最後に、全ての情報源を用いて、コンピュータによって、診断情報の最もよいセットをもつ各々の診断症例の複合症例記録が作成される。さらに、複合情報の一致性の確認が行われる（例えば、部位と組織型とに矛盾がないか、死亡日が最終確認日以前ではないか、など）。

症例決定システムは、ずっと、改善しつづけている。規則で決定できない症例や、矛盾のある症例は、OCR職員が調べ、コンピュータ規則を修正する場合がある。規則の変更は、システムに組み込むことができ、新しい規則に従って全登録ファイルの再加工ができる。

オンタリオがん登録での項目

3種類の記録が、登録室に存在する。即ち、元データ、複合同定記録、複合症例記録である。元データは、前述した外部情報源から入手し、がんに関連する出来事を表す。複合同定記録は、データ連結によって作成され、各がん患者に関する最もよい同定情報と人口統計学的情報とを表す。最後に、複合症例記録は、症例決定システムで作成され、各々の複合記録が、1つの原発がん症例を記述する。ほとんどの解析には、表2に示されている複合記録の項目を用いる。概して、腫瘍に特有なその他の項目、例えば疾患の拡がりなど、は元データで利用することができず、従って、OCRには含まれていない。

オンタリオがん登録の利点と限界

オンタリオの独特の登録システムの利点がいくつかある。複数の情報源が、質と完全性の高い罹患データを提供するために、まとめられている。近年の研究では、1982年のがん登録の完全性は、捕獲-再捕獲方法論を用いると、全がんで95%以上、最も低い皮膚の悪性黒色腫で91%、最も高い deep-seated（体表から到達しにくい部位）の消化器がんで98%以上であった(Robles et al., 1988)。ひとつの診断に関して複数の情報源から情報があれば、すでにわかっている各々の情報源の利点に応じて、各々の項目に最もよい情報を選択することができる。

複数の情報源は、ひとつの「患者のプロフィール」を作成し、がんの診断に関係する全退院症例のデータがOCRに保管されている。従って、入院期間やその他の保健医療計画策定者に役に立つデータが利用できる。患者のプロフィールを用いれば、同一個人における多重がんの同定も、簡単に行うことができる。

この独特の登録方法は、比較的费用がかからず、このことは、オンタリオの大きさの管轄区域では、重要な特色となる。

がん死亡と同様に、がん以外の死亡も、定期的に患者記録と連結されている。さらに、

表2. オンタリオがん登録室の複合記録にある項目

記録	項目
複合同定記録	登録番号 氏名 (alternates を含む) 性別、生年月日、出生場所 病院コードおよび住所コード 最終確認日 最終確認日の生死状況 もし死亡していれば、死亡診断書の死因
複合症例記録	がんの部位および組織コード 診断日 診断方法 診断時の住所 最も早い治療日 病院およびRCC/PMH ^a カルテ番号 診断した病院

a : RCC, 地域がんセンター; PMH, マーガレット王女病院

1964-1984年に診断された症例に関しては、オンタリオ運転免許証ファイルとの連結により、死亡せず、医療も受けていない多くの症例に関して、生存の同定が行われている。この2つの連結により、OCRの質が改善され（例えば、生年月日および居住地）、年齢、性別、部位別の生存率の解析を行うことができる。

最後に、コンピュータによる連結と症例決定システムとを使用することにより、一定の方法によってデータが加工されることが、保証されている。もし原発部位の割当方法を強化すれば、システムの改善によって登録データベース全体（50万以上の症例に関する600万以上の記録）が加工されるので、OCRの全期間の罹患データの質が向上しうる。さらに、多重がんの定義が変更したことによる罹患率への影響は、2つの症例決定システムを全登録データに適用し、結果を比較することにより、評価することができる。

しかし、システムの限界も、いくつか存在する。これらは、主に、入力データの種類、質、および流れの信頼性に係り、従って、制御にも限界がある。いくつかの情報源に関して、部位、組織型、居住地のコーディングは、病院の病歴室（病院退院記録の場合）や Office of the Registrar General（死亡診断書の場合）によって、登録室の外部で行われている。従って、OCRはこれらの情報源からのデータの質を、コーディングに関しても、受け取る記録数に関しても、制御することができない。それでも、州の約50%のがん症例が、RCCかPMHに紹介されており、そこではデータの質とコーディングの統一性が確認されている。

入力データ源のいかなる変更も登録に影響しうるので、がん罹患の時間的傾向は、情報源における変更や問題に関係した人為的なものである可能性がある。例えば、病院退院記録をMOHに提出する計画に関する1978年の行政的改革のために、OCRが受け取る退院記録の数は増加し、罹患率が突然上昇した。また、1970年代を通じて、自主的にOCRに提出される病理記録の数が劇的に増加した。従って、この期間には、OCRは、病理検査室からのみ報告された患者で、病院に入院せず、PMHやRCCsにも紹介されなかった患者の数が増加した可能性のあることを危惧している。特に通常入院の対象とならない部位に関しては、この期間を通じて、人為的な増加が著明かもしれない。そのような影響は、OCRが成熟するにつれて、頻度が減っていくであろう。

外部の情報源に頼っているもうひとつの問題点は、罹患データの作成が、結果として遅延することである。HMRIは、会計年度1年分の全ての退院記録を、年度の終わりに加工する。その後データをOCRに渡す。退院記録は情報源の大部分を占めるので、これらのデータが厚くまでデータ連結を行うことができない。1年分のデータを追加するためには、OCRで9か月の処理加工が必要となる。従って、ある年の罹患データが利用できるまでには、年の終わりにから18-24か月かかる。

さらに生じる問題点は、がん登録以外の目的で作成されたデータを利用することから生じるのであるが、完全な人口統計学的、地域的、腫瘍特有的情報が欠けていることである。例えば、RCCsもしくはPMHで診察を受けた一部分の患者を除いて、診断時の臨床進行度に関するデータが利用できない。以前は多くの患者に関して、現在は病理記録のみからなる患者に関して、居住地が不明のまま残る。年齢と正確な生年月日は、重要な連結項目であるが、特に病理報告で、時折、記載されていないことがある。これによって、年齢不詳で届けられる症例が、比較的高い割合（0.06%）を占めている。

さらに、OCRでは、いくらかの過剰報告がある。大部分の重複記録（同一個人の記録が連結過程でまとめられない）は、NYSIIS以外のポケットを用いた内部連結過程で除かれているが、まだ0.2%の重複登録があると推定される。これは、個人が固有の識別番号をもっていない州では、解決不能な問題である。しかし、全記録の正しい連結ができなかったための、過剰報告の問題の大きさは、完全な手作業による連結に比べると、格段に小さいであろう。手作業で連結をおこなっていた1965-66年の罹患データに現行のシステムを適用して比較すると（MacKay & Sellers, 1970, 1973）、症例数が11%減少した。

結論

現行のOCRは、新しい登録室である。なぜなら、1972-76年の罹患データは1983年に作成され、1977-1982年のそれは1984年に作成されたからである。しかし、いまや登録技術がよく確立されたので、罹患データは毎年追加され、年の終わりの18-24か月後に利用することができ

る。この期間を短くする努力は、病院記録や病理記録のコンピュータ化の増加とともに、持続して行われており、より短い間隔で罹患データの作成が可能となるであろう。OCRは、その他の定常データ源をふやすことによって、OCRの完全性と質とを改善するために、例えば、細胞診記録、血液診記録等の、新しいデータ源を常に調査している。さらに、連結システムおよび症例決定システムは、持続的に改善され、近代的に効率化されており、それに応じて、OCRの罹患データの質と計測までのタイミングが改善されるであろう。OCRの洗練されたコンピュータによるデータ連結能力によって、健康システム以外の情報（Ontario運転免許証ファイルなど）を、人口統計的項目と最近の状況とを更新させるために、OCRに連結することができる。

コンピュータ技術を使った、革新的な方法によるがん登録は、組織学的確診率が85%を越え、死亡票のみによる登録率が約2%であるような、質のよいがん登録室を生じさせた。健康システムの種々の局面でコンピュータが導入されていくに従って、OCRのコンピュータ・ベースのアプローチは、未来のがん登録の最適な方法であることが証明されるであろう。

（翻訳：味木和喜子、花井 彩）

Appedix 3(d) The Department of Helth-Rizal Cancer Registry

A. V. Laudical and D. Esteban

付録3（d）衛生部ーリザールがん登録室 （フィリピン）

〔 省 略 〕

付録4. CANREG： マイクロコンピュータ用がん登録ソフトウェア

M. P. Coleman and C. A. Bieber

緒言

がん登録データの蓄積、処理、解析のためにコンピュータは今日広く利用されており、利用されているハードウェア、ソフトウェアの範囲も相当なものである。世界のあらゆる規模、すべての地域の61のがん登録室におけるコンピュータの利用に関する調査で、ハードウェアメーカーについては20種類、データベース用ソフトウェアでは12種類、コンピュータ言語と統計パッケージでは30種類が利用されていた (Menck, 1986)。

マイクロコンピュータを利用しているのは、24登録室だけで、主に、あるいはマイクロコンピュータのみを利用しているのは、5登録室のみであった。

しかしながら、多くの小規模登録室、あるいは最近開始した登録室では、そして特に開発途上国では、コンピュータによるデータ入力と処理にあたり、マイクロコンピュータだけが唯一、実現可能な選択である。マイクロコンピュータは安価で操作が簡単で、メンテナンスもあまり要らないうえに、ほとんどの場合、相互の共通の互換性がある。また登録室の他の機能、例えば文書処理などの市販ソフトを広く利用できる。

本章は、マイクロコンピュータを利用した、がん登録用のソフトウェアパッケージであるCANREGについての、簡単な記述である。CANREGは、非常に限られた予算とコンピュータの専門的知識が少ないか、または全くなく、比較的少ないデータ量（データ項目が最大45項目で、年間約3000件までの症例数）のがん登録のために、特に設計されたものである。また、低学歴（ほぼ6年～10年の就学歴）のがん登録室員が使えるように設計されている。CANREGは、もっと強力なソフトパッケージが利用できるミニコンやメインフレームには適さない。米国の大きい登録室は50万米ドル以上の費用をソフトの開発と検証にかけている。一方、CANREGは、6000ドル程度の少ない費用で導入できる。

CANREGの主な特徴

CANREGは、がん患者についての個人的、人口学的、腫瘍関連情報をコンピュータに入力し、データの完全性と妥当性について自動的にチェックし、検査、修正したあと、最終的にリスト、頻度表、がん罹患統計を作成する目的に利用できる。

プログラムはメニューによって操作でき、ごくわずかしか訓練を受けていない者にも操作できる。システムはユーザーにやさしく、柔軟性に富み、開発途上の登録室の要求に応じて繰返し変更が可能である。

入力データ (画面)		
登録番号	住所	組織診
部位	生年月日	診断根拠
病院番号	年齢	死亡日
多重がん	性	剖検
氏名	診断日	

図1. CANREGによるデータ入力時の画面の一例—フィジーがん登録室

注：名前とデータとは作りものである

個々の症例のデータの入力

各がん症例のデータは、通常は、キーボードで入力され、その際、各登録室の必要性に応じようにデザインされたデータ入力画面が用いられる。(図1参照)

データ項目の多くは、コード形式で入力され、各データ項目について可能な限りの有効入力数値(例 1=男性、2=女性、9=不明)を保有した辞書を参照して、自動的にチェックされる。データ入力時に、コードに対応する内容が画面に表示され、操作者は原データが正しくコードされたかどうかを確認することができる。例えば、形態(組織型)について80503を入力すれば乳頭状癌、NOSが表示され、記載記録と比較できる。当該項目の辞書に存在しないコードは拒絶される。

30変数までは、このように辞書内コードとして保有でき、CANREGシステムではICD-0の局在と形態学(部位と病理の)コードの辞書が、英語もしくは仏語で提供されている。辞書は簡単に改変でき、新しい辞書が作られる。従って、対象地域における民族グループ名を各々の民族の言語で辞書に含めることができる。同じようにして、病院、検査所、及びその他のデータ源が別々の辞書を用いてコード化できる。

架空患者についての単純なデータ入力画面を、図1に示す。データ入力者が部位について1429と入力し、対応する局在(部位)が次に表示されている。病院についてコード2と入力されて、対応する病院名が適切に辞書と対応して表示されている。データ入力の辞書制御は、登録データの内部妥当性を維持する重要な方法である。

患者名、誕生日、腫瘍部位など、いくつかの項目は、どの登録記録でも必須である。つまり、プログラムは、新しいレコードがデータベースに追加される前に、これらの項目の欠損がないかどうか確かめる。どの項目が必須情報であるかの選択は、ユーザーによって定義され、後に改変もできる。

表1 CANREGのデータ入力時の内部ロジックチェック (a)

- 1 登録番号が未だ割付けられていない
- 2 性と部位の妥当性
- 3 日付の妥当性、ありうる暦日
- 4 関連した日付の妥当性
 4. 1 誕生日が診断日、死亡日より以前である
 4. 2 診断日が死亡日より後にこない
 4. 3 診断年、死亡年が現在年より後にこない
 4. 4 年齢が、誕生日と診断日とから計算した年齢に矛盾しない
- 5 基本項目について欠損値がない
- 6 辞書保有項目に対してのコードが妥当
- 7 数値項目が英字コードを含まない
- 8 オプション

注a 5から8までのチェックはユーザーが定義可能

各レコードの内部ロジックは、データ入力時のいくつかのチェックの際にも行われる（表1参照）。どのチェックでも、はじきだされたレコードはデータベースに記録されず、適切な修正を促すメッセージが表示される。

データ入力処理におけるこれらのチェックは、データベースに加えられるすべての腫瘍レコードが、完全に妥当であることを、一貫した基準で確かめることによって、登録データの質を高めるように設計されている。

コード化データの使用

CANREGにおいては、ほとんどのデータをコード形式で入力、保存するようにした。これには、いくつかの理由がある。

(1) 保存効率

ほとんどの場合、コードを使用するほうが記述情報を使用するよりも、保存にあたり領域を占有しない。病理形態のようなデータ項目については、コード化することによって得られる保存スペースの節約は相当なものとなる。（例：44文字に対して5桁）

(2) 入力速度

コード入力は文書入力より速く、またエラーがおこりにくい。データ入力に先だち、コードを手作業で与える必要があるが、それ故データのコード化は、レコードの妥当性チェックにおける重要な段階である。

(3) 妥当性のチェック

データが数値化されていると、妥当性チェックが容易である。

(4) 言語互換性

局在と形態（部位と病理）のコードの使用は、いわば、英語版のCANREGで作成されたデータを、仏語版で改変することなく、解釈し分析することを可能にする。なぜなら5つ、ないし6つのまれに使われる修正コードの例外はあるものの、双方の辞書が数値的に同じであるためである。

データ入力：ファイルのとりこみ

コンピュータ上の既存のデータファイルは、標準（ASCII）形式であればCANREGにとりこめる。データレコードの項目順序は、CANREGと同じでなくてもよい。個々のレコード入力時になされるロジックチェックは、しかしながら、とりこまれたデータセットに対しては行われない。

データマネージメント

データベースの腫瘍記録は、登録番号を入力することにより、画面に呼び出すことができる。その際に記録は、例えば死亡日や改訂された組織診報告について、更新したり、修正したりすることができる。前述した内部ロジックチェックは、レコード変更時においても、データベースが更新される前に、実行される。レコードはこの段階で、削除することもできる。

データの保護

プログラムはデータベース全体に関するパスワード保護を組み込んでいる。プログラムそのものは、コンパイルされた形で提供されており、データベースの解読を即座にすることはできない。多くの項目をコード化しておくこと、さらに有利な点としては、データレコードが概ね数値形式であることである。もしもマスターのデータファイルを含んだフロッピーが郵送中に紛失したとしても、関連辞書がなく、データ構造を定義するファイルがなければ、解読はできないだろう。

プログラムには、データベースのバックアップコピーをフロッピーディスクに作成する機能があり、第一の保管場所として使われるハードディスクが故障した場合に、バックアップディスクからデータベースを再構成できる機能をもっている。毎週間隔で、あるいは大容量のデータであればもう少し頻回に、ユーザーはバックアップするように勧告される。

サブセット、一覧、表

CANREGシステムでは、一度に3つまでの項目（例 部位、性、年齢）を用いて、例えば、

40-69歳の男性の全肺がんを含むファイルを作成することができる。抽出は、オリジナルのサブセットに繰返し行うことができ、別の項目を用いて、例えば組織診で定義された、より特異的なサブセットを抽出することができる。

個人データの一覧は、印刷することも画面に表示することもでき、出力の形式は、自由にユーザーが定義、変更できる。一覧は、他のプログラムで加工できるように標準（ASCII）ファイルに出力することもできる。一覧は順序を指定（例えば部位、性、名前の順）でき、特別なタイプ（例えば男性のみ）のレコード、あるいは値の範囲（例えば部位コード150-159）に限定することができる。

図2に示す例は、部位順の患者一覧（架空）を示す。この例で、部位、組織コードがこれに対応する記述とともに、印刷時にコード辞書から自動的に抽出され、印字される。選ばれた事例は、中規模登録でしばしば使われる一覧である。すなわち、二重登録の探索の手法として、性、名前ですらにソートすることができる。

登録患者リスト							
登録番号、	氏名、	年令、	性、	部位コード、	同記述、	組織コード、	同記述
”	”	”	”	”	”	”	”
”	”	”	”	”	”	”	”

図2. CANREGによる登録患者リストの一例

訳者注：原画面では、患者1人のデータが横1行に示されており、縦にならんで出力される。

データ解析

ほとんどの登録室は、例えば性、年齢、部位別に登録された症例数の年毎の表、あるいは年齢別罹患率に関する同様の表のごとく、標準的な表を定期的作成しなければならない。

一連の解析プログラムは、それ故、CANREGに組み込まれている。解析プログラムは、前述したCANREGのメインシステムとは分かれているので、別のコンピュータシステムで作成されたデータの解析もできる。解析プログラムは、CANREGが出力ファイルとして事前に作成した形式のデータを扱うように設計されている。（以下を参照）

主要解析プログラム（CRG Table）は、性（男性、女性、合計）別、年齢階級別、がんの部位とタイプ別の患者数を表にする。年齢階級はユーザーが選択でき、がんの部位とタイプはCANREGのデータに入力されたものである。表は、年齢別の総数（年齢不詳のカテゴリーを含む）と部位別の総数を含み、また全がんに対する各々のがんの頻度（百分率）を含む。もし人口データが提供されれば、プログラムは年齢、性、部位別の罹患率表を作成し、粗率、年

年齢調整（世界人口）率の表を作成する（図3参照）。小児腫瘍について用いられる診断群分類は、ICD-0の部位と組織とのコードから自動的にひきだされ（Birch&Marsden, 1987）、そして年齢階級は、0歳、1-14歳、10-14歳である。

	年齢階級別 罹患率	全年令
部位	全 0- 5- 70-	粗率 調整率
・ ・ ・		

図3. CANREGによる作表例-部位別、性別、年齢階級別罹患率

CANREGの他の機能

CANREGシステムは、がん登録に関連するいくつかの追加機能をもつ他のプログラムを組み込んでいる。例えば

- オペレーターが個々のICDコードの妥当性をチェックできること、そして英語、仏語（ICD-9とICD-0）、スペイン語（ICD-0のみ）で、それぞれのコードの定義をテキストで表示できること。
- ファイル中のICD-0コードを、対応するICD-9コードに変換すること。
- ICD-0の部位と組織とのコードを使って、小児腫瘍の群、コードをデータファイルに作成し、年齢、性、腫瘍型別に症例数と率の表を作成すること。

データセットの出力

CANREG解析プログラムやその他のソフトウェアを使って、協同研究のためあるいはローカルの解析のために、他の登録センターにデータを渡す場合には、CANREGデータベースは、標準（ASCII）ファイル形式で出力することができる。

必要なハードウェア

CANREGを実行するために必要なハードウェアの基本仕様は、次の通り。

- ディスプレイとキーボードをもち、IBMPCと100%互換性のあるマイクロコンピュータで、最低 256kRAM実装、10MBのハードディスクとフロッピーディスクドライブ装置が

必要である。DOSはバージョン 2.1以降のPC-DOSかMS-DOSが必要である。

－ ドットマトリックス・プリンター

もしコンピュータを用いて、その他のアプリケーション・ソフトも利用するのであれば、RAMは512か640kでハードディスクは20MB以上が望ましい。ハードディスクの故障の場合のデータの安全性のため、バックアップ用のテープストリーマーを備えることも考えるべきである。

電源供給が不安定で、雷に対する配電システムの保護が適切でない国では、安定化電源装置がコンピュータの電力供給には適当であろう。除湿器と空調設備は、極端な湿度と温度に対処するために必要になってくるかもしれない。

基本的なコンピュータ装置の価格は、国によって異なっているが、たいてい 6000米ドル以下である。操作にあたっての費用は、仕事量によっても幅があるが、以下に示す項目について検討しておくことが必要である。

消耗品

- プリンター用紙
- プリンターのリボン(もしくはインクなど)
- フロッピーディスク
- 電源
- テープ (ストリーマーバックアップ用=オプション)

オプションハードウェア

- テープストリーマー装置
- 安定化電源装置

その他

- ハードウェアの保守契約

導入、解説書、職員の訓練

CANREGの導入には、登録室の要求にあわせて正確な定義をする必要がある。これらの要求は CANREGの改変版に組み込まれる。改変版は、該当登録室にあわせた特別なものであるが、なお、他の CANREGシステムと互換性がある。CANREGのひとつの設計基準は、プログラミングの技術が全くないユーザーの要求にあわせて、改変することを可能にすべきであるという点である。このためには、収集するデータ項目を定義するデータ入力用画面や、出力形式の設計、種々の辞書の構築と改変、そして実データを用いてシステム全体を検証するために、数日間の初期投資を必要とする。システムの各々の要素は、その結果、ユーザーによって改変できるようになる。

完全なユーザーガイドが提供されており、データ入力事務員は現地で訓練される。操作の初期段階で発生する問題解決にあたり、また解析やデータの提示を含むがん登録の他の側面

についても助言するため、支援サービスが提供されている。

導入したユーザー

1990年初頭までにCANREGは12以上のがん登録室で導入し使用されているが、主にアフリカ、アジア、ラテンアメリカやオセアニアなどの開発途上国である。

CANREGの限界と受注開発

CANREGシステムは、マイクロコンピュータを使用することを考えて特に設計されており、メインフレーム上では使用できない。データ量が非常に大きい場合、又は、1レコードあたり45を越えるデータ項目をもつ場合も適当でない。

システムは継続して検討が加えられており、以下のような多くの改変が、進行中、もしくは検討中である。

- 個々のレコード入力時におけるデータ検証の追加。例えば部位と組織型との組み合わせがありそうにない、もしくはあり得ない場合の警告や拒絶。
- 取り込もうとしているデータファイルに対して、個々の症例を入力する時に使ったのと同じ妥当性チェックを行うオプション。
- 全データセットにおける重複登録の可能性を検索する場合に用いる一群の項目（例えば、名前、誕生日、腫瘍部位）を定義するオプション。
- 単純化された、ユーザーが会話的に導入できる手法。
- データセットの選択的（すべてのデータ、あるいは最後のバックアップから追加したか改変したレコードのみ）、あるいは辞書、索引に関する選択的バックアップ機能。
- 出力形式の設計や改変における簡単な手続き。
- 解析や表作成機能の拡張、例えば、ドットマトリックス・プリンターのみを使用して年齢別罹患率曲線を作図すること、など。

CANREGシステムの入手

システムは、IACRの登録メンバー、もしくはIACRと共同している登録室であれば利用できる。開発途上国の登録室に対しては、現地でのプログラムの改変のため（上記参照）、またコードや辞書、及びシステムの他の要素が、ユーザーの要求に合致することを保証するため、さらにシステムを導入し、それを利用するスタッフを訓練するために、支援が供与される。

詳細は、次のところに問い合わせ願いたい。

Chief, Unit of Descriptive Epidemiology, International Agency for Research
on Cancer, 150 cours Albert Thomas, 69372 Lyon, Cedex 08, France.

(翻訳：陶山昭彦、津熊秀明)

文献（10頁）および索引（8頁）

〔 省 略 〕

Cancer Registration—Principles and Methods— (訳)
がん登録の原理と方法

平成8年3月 (非売品)

- 発行 厚生省がん研究助成金 (5-3)
「地域がん登録の精度向上と活用」に関する研究班
大阪市東成区中道1-3-3 (〒537)
大阪府立成人病センター内
- 協力 地域がん登録全国協議会
(大阪府立成人病センター内)
- 印刷 久山日進社
大阪市東住吉区杭全4-7-1

