

疾病構造解析システム (SAGE) の開発と運用

(財) 東京都臨床医学総合研究所

倉科周介

1. はじめに

わが国の疾病構造は、第二次大戦後から今日に至るまでの間に著しく変貌した。死因統計について死亡者の年齢分布を経年的に観察すればこの状況は如実に明らかとなる。すなわち1955年頃までは、乳幼児死亡が極めて多く、また青年期においても著明なピークが形成されていた。この初期死および途中死のピークは以後20年余の経過を通じて迅速確実な減少を示している。その結果として日本国民のいわゆる平均寿命が延長し、世界有数の長寿国が現出するに至った。

初期死と途中死の減少に寄与したものが、1940年代に登場した抗菌性抗生物質であったことは明らかである。感染症による死亡の減少は抗生物質の効果のみでは説明困難とされているが、他の要因についての分析は今後に残された課題となっている。

一方、死因としての感染症の比重が極度に減少するに伴って、非感染性疾患が時代の脚光を浴びる情勢になった。死因についても、悪性新生物、脳血管疾患、心疾患、糖尿病、肝硬変などが重要とみられるに至っており、これらの疾患に対する抜本的な対策が急務とされている。

上記の如き重要な非感染性疾患は、一般に直接的な病因を特定することが困難で、発病後の経過が慢性であるのみならず、患者体内における病的機転の作動から臨床的な発病に至るまでの潜伏期も長期に及ぶことが知られている。すなわち長潜伏期型多要因性慢性疾患という特性を持つ。さらにまた、臨床的発病をみた後では、これらの疾患に対する処置はしばしば姑息的にとどまらざるを得ない。すなわち理想的な治癒は不可能に近い。したがって疾患の全経過にわたる特性を個人および集団のレベルで的確に把握し、標的を明確にする努力こそが必要である。有効な対策の開発は、こうした知識基盤の構築が前提となっていて、はじめて可能になるものと考ええる。

2. 最近の疾病構造の変貌における若干の特徴

難治性の疾患の場合、その転帰は多く死亡である。疾病は連続性と反復性を有するため定量的観測には多くの技術的困難を伴うが、死亡という事象は生物の個体において、出生とならんで数少ない単独生起型事象である。したがって死亡現象の発生状況の観測は、疾病構造の変貌を検討する上で有力な手掛りを与えるものと期待される。

こうした着想に基き、1955年以降の人口動態統計死因統計のデータを系統的に分析した結果、極めて興味ある結果を得た。すなわち、肝硬変による死亡は経年的に増加の傾向に

あるが、特に男子の死亡の増加が著しい。しかも死亡者の年齢分布のパターンは、当初、60歳代を中心とする単峰性であったものが、年次の経過に伴って、次第に若年側に新たなピークが生じ、これが成長しつつ高年側に移動するという変化を示した。そして最近（1975年）では、45歳と65歳にピークを持つ二峰性分布のパターンとなった。しかも、この分布は遅れて出現した若年側ピークが、従来の高年側ピークをはるかに凌ぐ数に達している。これに対して女子の肝硬変死亡の年齢分布パターンは、同じ25年間を通じてほとんど不変であった。男子の肝硬変死亡に見られるこの特徴を説明するため、いわゆる出生コホートの概念を導入して、分布パターンを分析すると、1930年（昭和5年）を中心として、その前後の5年ほどの期間に出生した男子世代が、上記の若年側ピークの形成に寄与していることが明らかになった¹⁾。

この知見を契機として、他の死因についても類似の世代的特性の有無を探索した結果、男子の自殺について、同様の世代特性が認められた。まず年齢分布のパターンは、当初、20歳を中心とする急峻なピークと、50～60歳における極めて平低なピークを持ち、いわゆる壮年期には陥没していた分布形態が、近年に至って大きく変化した。20歳のピークは著しく縮小してむしろ25歳の位置に移行し、一方、従来の40～45歳における陥没部が増大して新たなピークを形成する傾向を示す。肝硬変と同じく世代的にみれば、当初の若年側ピークを形成したのは1930～35年の出生世代であり、最近の中年期ピークに寄与しているのも、また同じ世代である。すなわち、この世代は終戦直後と最近との2回にわたって自殺の多発を経験するという極めて特異な集団である²⁾。

この世代の男子における最近の肝硬変および自殺による死亡の増加傾向は、現在、変化の途上にある。したがって、この世代の構成員全体における両死因の比重がどのようになるかは、世代の集団としての生命がほぼ完結するとみられる約50年の将来において、ようやく明らかになるものと考えられる。しかしその将来を待つまでもなく、死亡パターンの特異な変貌から、この世代集団の男子が、それに前後する出生世代に比して著しく特異な反応特性を内蔵していると推測することは可能であろう。

なお、この世代の男子については、糖尿病、原発性肝癌などによる死亡の増加を示唆する知見もある。さらに観測と分析を拡充して追跡する必要があるものと考えている³⁾。

3. 非感染性慢性疾患の病態的特性とSAGE開発の契機

極めて素朴に疾病現象を考えれば、疾病の契機は宿主と病因の接触によって形成される。この際、疾病機転が宿主の体内で発動されて、進行を開始するか否かは、病因に対する宿主の反応特性、すなわち抵抗力もしくは感受性と、作用要因としての病因の特性、すなわち毒力との相対的關係によって決まる。そして疾病機転が進行する過程で、さらに種々の要因が作用するに伴って、宿主の反応特性が連鎖的に変化して行く。

このような関係は、作用要因を病因のみに特定することなく、ひろく生物の個体と外界

の事象との間に一般的に存在する関係に拡張して考えることが可能である。すなわち個体の生涯のある時点において、外界から作用した要因に対する個体の反応特性は、それまでの生涯で個体が通過してきた経路のひろい意味での作用特性に依存すると考えるものである。

個人の反応特性は、出生時のいわば先天的な状態を初期値として、以上のような過程によって形成され、変化して行く。これが発育を含めた加齢の過程である。疾病はこの過程における特殊な反応であり、発病以前の反応特性と病因の作用特性の相互関係からもたらされる結果であると同時に、以後の反応特性を変化させる要因の一部ともなる。

個人もしくはその集団が、生涯において通過して行く時空間世界の経路上の作用特性は位置によって異なる。そして、この作用特性が質的あるいは量的に他と著しく異なる位置を通過した個人や集団は、その位置を通過しなかった人々に比べて異なった反応特性を獲得する可能性がある。このような反応特性の修飾、およびその結果としての特異な反応の発現を、経路効果 (route effect) と呼ぶことを先にわれわれは提案している³⁾。

非感染性慢性疾患の病像が、このような経路効果によって規定されることは、十分に推測が可能であり、そのような観点から疾病現象を観測することは、学術的ならびに実務的に多大な意義があろうと思われる。SAGE 開発に着手した契機はここに存する。

4. SAGE の基本構築

後段に詳述するように、SAGE は最終的には多目的型の知識生成システムを指向している。しかし、ここでは直接目標である疾病構造解析システム (System for Analysis of Generalized Epidemiology) との関連において、開発思想、構造および機能の概要を述べる⁴⁾。

A. 開発思想

前述のごとく、疾病現象は生物学的連続性と多次元的展開を持った存在であり、それに関する認識は観測行為の特性に規定される所が大きい。ここで特性というのは観測者、観測対象、観測手段の特性のほか、観測者と観測対象との位置関係をも含めた広義のものである。静止した事象や実験室科学の研究対象に関する認識は、観測の反復によって当初の認識産物における欠落の補填や誤謬の修正が可能であるが、ヒトの疾病現象は本質的に野外的存在であり、動的かつ個別性が強いいため、原則として観測行為は反復不能である。したがって時空間世界の中における疾病現象の展開を把握するためには、特性の異なる観測行為の成果であるデータを、疾病現象の生物学的連続性に即して知識装置上の仮想空間内に構造化し、これを多面的、多角的に観測、分析することにより、直接観測データに存在する存在制約性 (Seinsverbindenheit) を解除してやるが必要になる。

B. データベースの構築手法⁵⁾

既存の各種観測記録から一次データとしての観測値、もしくはそれに最も近いと評価さ

れる二次データのみを抽出し、観測対象の一般的、普遍的な属性に基いて、層別化、区分化して蓄積する。これは記録の編集特性に発するデータ集団の特殊構造を一旦解体する作業であり、脱構造化（destructurarization）と呼ぶ。また、記録の構成要素である原データを基本的な処理対象とするという意味において、要素化（elementalization）とも呼べる。

この要素化されたデータ集団をもとにして、過去に存在し、生起したと考えられる事象の像を仮想空間内に再構成する（real world reconstruction）。ここに実現されるデータ構造をreal world databaseと呼ぶ。

C. SAGEのデータベース構造

疾病現象の発現基盤は人口集団とそれが構成する社会の構造にある。わが国は国家形成の過程を通じて、若干の例外的事象を除き、国際関係の面で極めて閉鎖度の高い社会システムを構築して現在に至っている。そのために国土、民族、言語、社会その他の国家システムの構成要素が時空間世界の中で占める物理的範囲の一致度が格段に高いという点で、わが国はいわゆる先進諸国の中でも特殊な位置にある。したがってこれら諸要素の相互関係を観測するには最適な対象ともいえる。

C-1. データベースの素材

国家システムを構成する諸要素に関する大規模かつ長期の観測指標のうち、とくに当面の研究目的に必要性の高いものから順次対象とする。

- (a) 人口現象データ：国勢調査、人口動態統計総覧など
- (b) 疾病現象データ：人口動態統計死因統計、患者調査、特殊疾患患者登録調査など
- (c) 社会現象データ：国民総生産、食料需給表、農業基礎統計、エネルギー統計、工業生産諸統計など

C-2. データベースの構成

(a) main frame database (MFDB)

疾病現象、人口現象、社会現象などの総枠の時空間的展開を表現するデータベースである。わが国の国家システムを構成する諸要素の表現は、この総枠を著しく越えるものではない。総人口、総出生、総死亡などがこれである。ただし、このデータベースを構築するにあたっては、時間軸を遡るにしたがって、総枠の規模を把握するための障害が出現する。特に第二次大戦終結前後の1944-46年の期間は各種の長期統計がほぼ例外なく欠損もしくは不備となっており、この間に発生した事件の精密な把握は不可能である。

また、それ以前の時期については、旧日本帝国の海外領土およびそれに準ずる地域に在住した日本人集団と、軍籍所属者集団の問題がある。前者については調査自体が本土のそれに比して著しく困難であったと推察されるが、ほぼ本土に準じた事業体系が維持されていた。国勢調査を例にとれば、1920年以降1935年まで4回にわたって本土と並行して調査が行なわれ類似の編集形態をもって記録されている。しかし、1940年については経済関係

の数値表が公表されたのみで、人口関係の諸値はついに公表されぬまま今日に至っている。したがって、1940年当時 250万名と推定される海外領土在住邦人と外国在住邦人75万名に関するデータは正確に把握することが困難である。

なお、陸海軍区画所在者、すなわち軍人および軍属については軍機上の問題があり、これもデータの把握が不可能に近い。敗戦時の兵員数は約 800万名といわれるが、これと外地居住者を合計すれば1000万名を越える。敗戦後の日本人口は復員および引揚により、この様な特異な生活経路を有する個体を大量に収容することからはじまっており、その規模は総人口の10% 前後に達したものとみられる。この部分に関する適切な推計を行っておくことは極めて重要と思われる。

(b) subset database

main frameを構成する現象諸要素に関するデータベースである。人口および疾病現象を例にとれば、総人口や総死亡に関する内訳（性別、年齢階級別、死因別、地域別、などの層別化、区域化データ）に基づくデータベースである。同一次元におけるデータの総計は、理論的にはmain frameのサイズに一致する。

ただし、このデータベースについては、層別化、区域化の基準が、調査年次および事業主体によって一定していない。また地域別データは、記録自体の保存状況が年次を遡るにつれて地域ごとに著しく異なる。昭和20年代以前については極めて非定式かつ断片的であろうと推定される。

(c) spot database

main frameの中で発現した事象の特定部分を対象とする非定型または不連続な観測によって得られたデータに基いて構成されるデータベースである。こうしたデータの発生母体となる事象は、一方でmain frameまたはsubsetのデータの構成成分ともなっているはずであり、したがって、データのサイズはmain frameやsubsetのそれを越えることはない。施設統計や個人資料がこれに属する。

D. データベースの構築と運用上における留意点

このシステムの当初目標は、わが国における疾病現象が時空間世界の中で示す構造を、現在にはじまって次第に過去に遡りつつ、知識装置上の仮想多次元空間内に再構成する所においた。

予備的検討の段階では、主として疾病分類の整合性の観点から1955-74年の期間を対象としたが、25年という時間は現在の日本人の平均寿命の約1/3に過ぎず、前述の経路効果を観測するには不十分である。

一方、人口現象および疾病現象に関係する官庁統計のデータ環境を見ると、人口動態統計の中央集査開始が1988年（明治32年）であり、国勢調査事業の開始が1920年（大正9年）である。したがって、これらのデータの活用により、疾病現象の構造解析の基盤となる人口現象および死亡現象のmain frameおよびsubsetに関するデータベースの構築は、少なく

とも1900年以降については可能と推定された。1900年出生の世代は今日において成員の大多数が生涯を完了している。すなわち、この世代はその生涯経路に関してほぼ完備したデータを保持して群の生命を完結しつつある、わが国最初の人口集団であるという点で、極めて重要な意義を持つ。

以上の理由により、SAGEの歴史的時空間軸の起点は1900年を選んだ。すなわちSAGEのMFDBの時空間域は、時間的に1900年以降、地理的空間は日本国領土内（戦前については海外領土を含む）、人口集団は国籍上の日本人とした。特に現段階では以下の諸点に注意をはらった。

- (1) 旧日本帝国海外領土及び陸海軍隊区画における関連諸データの蒐集
- (2) 1944-46年を中心とする終戦前後の欠落データの補填
- (3) ICD改訂に伴う疾病現象の統計的不連続性の補正
- (4) 個体の生涯の連続性を考慮した人口現象及び疾病現象の遡及的コホート観測

5. 経路効果に関する基礎的検討

物理学的に定義される時空間が、生物の個体が生存する空間（領域）を表現していることは確実である。さらに個体の出生から死亡（参入から退出）までの経過も、この空間内の軌跡として表現されるように見える。しかしこの軌跡が特異性を持った個体の移動を表わし得るには、上記の空間が存在するのみでは不充分であり、軌跡の始点の位置情報が組合わされて、はじめて個体の時間的特性としての年齢が定義可能となる。

物理的な時刻そのものは個体の年齢を示すものではないが、物理的時間の進度（時間経過速度）はこれによって慣習的に計測されているため、集団の成員について共通と仮定する。こう仮定すれば、年齢は（また物理的時間も）この速度の積分値であるため、これを当該個体にとって一意とするには、初期値（すなわち積分定数）を与えねばならない。にも拘らず、従来はこのパラメータが明確に分離して取扱われることがなかった。これを独立させることにより、はじめて集団内における個体を明確に特徴づける真の時空間が認識可能となる。このパラメータを個体年齢とよぶ。

ところで、上述の物理学的時空間と個体年齢とを組合せた空間では、個体の存在経路を記述する空間としては普遍性を持たない。この空間では個体年齢が時間軸と独立でなく、その内部に写し込まれており、個体年齢が変化すれば時刻も変化してしまうからである。

そこで個体の集合があらゆる始点を含むものとして、この集合を対象とすれば、個人の年齢経過でなく、集団の年齢構成が時間軸上の時間の流れと独立して存在することになる。このことから、物理学的時間軸と集団を構成する年齢軸とによって張られる平面が、集団としての個体が存在する普遍的時間をなすと考えられる。ここで物理学的時間軸を暦年軸、以上の意味の年齢を示す軸を年齢軸、この両軸によって張られる平面を暦年-年齢平面とよぶ。またこの平面と物理学的空間軸とが構成する超空間を集団的時空間または真の時空間

間とよぶことにする。

以下、議論は暦年－年齢平面上に限定する。先の仮定にしたがって、個体の物理学的時間の進行速度は、暦年軸上でも年齢軸上でも、その写影は同一で、同じ単位で計測できるとする。

真の時空間内における個体の軌跡のprojectionを考える。暦年－年齢平面上で各々の個体の軌跡はかならず暦年軸上に始点を持つ傾き1の直線となる。すなわち

$$L = \{ (b + t, t) \mid t: \text{個体年齢}, b \in \text{暦年} \}$$

ただし、 L と暦年軸との交点を $(b, 0)$ とし、この点において $t = 0$ とする。ここで b は出生年である。 L を個体生涯直線とよぶ。 L 上に個体年齢 t が対応させられているというまでもないが、年齢軸に L を正写影すれば、対応する点で t と年齢が一致する。

この L と少なくとも一部が重なる個体生涯直線を持つ個体の集合を世代 (generation) とよべば、これらはすべて同じ暦年軸切片 (出生年) を持つ。よって世代は出生年すなわち暦年と1対1に対応している。また、ここで個体の集合を世代と考える代りに、世代の確率的な一表現が個体であるとも考えることも可能である。個体の軌跡に対して、世代の軌跡を世代生涯直線とよぶ。

個体 (世代) 生涯直線は、個体年齢とともにある個体 (世代) の時間的経過を表現しているものである。よってこの直線上の各点に個体の持つさまざまな特性を対応づけることができる。特性を表わす何らかの集合を特性空間とよべば、これは生涯直線からの写像の値域である。また、この写像は各々の生涯直線によって、その形が異なるであろうと考えられる。

具体的な特性空間として、最も簡単なのは、個体の時間経過、すなわち個体年齢である。これは従来あまり明確に認識されていなかったという点で特に重要であろう。その変形としての個体の生理的あるいは心理的な時間経過がある。これは肉体的ならびに精神的な加齢過程の表現と考えてよい。個体における疾病の自然史的な経過についても、同様な考えに基いての処理が可能である。

6. 考察

ヒトの疾病、特に長い時間的経過を持つ慢性疾患の全体像を把握するためには、世代生涯直線に沿った長期の観測が必要となる。この種の観測もしくは研究の手法はコホート研究 (cohort study) とよばれる。これにはprospective study とretrospective study がある。研究着手の時点において、一次データが時間軸に対して順行的に発生するか、遡及的に発生するかによる違いである。特に後者についていえば、これは他の観測者の認識産物の利用の上に立つ研究である。わけでも人口現象に関するデータの如きは、長期に及ぶ大規模な行政的統計調査事業の所産であり、その意味では、先人が企画実施したprospective studyの成果があつて、はじめて研究が成立するとさえいえる。ただこの種のデータ

は事業独自の観測および記録に関する特性があり、retrospective study においては、研究目的に沿った再編成が必要となる。この点についてはCaseの詳細な検討があり、著者らも若干の提案を行ってきた^{6,7)}。

しかしコホート特性あるいはコホート効果は、先に述べた経路効果の部分現象にすぎない。より本質的なのは、個体や集団がその生涯で通過する多次元空間内の経路の特性である。この特性は厳密に言えば、通過者に対する経路側の作用特性と、通過者側の反応特性とに分かれ、両者の相対的關係によって規定されるものである。すなわち通過点の超空間内位置と、通過時における個体もしくは集団の生涯直線上の位置によって決定される。

このように考えれば、特に野外事象の観測値に位置表示情報を付けることの重要性は明らかであろう。適切な位置情報を持った観測値によってのみ、認識空間内におけるreal worldの再構成は可能だからである。また、こうした条件を満たすデータ（仮にpositioned dataとよんでおく）こそがreal world databaseの構築に役立つ認識産物であろう。positioned dataはヒトの活動に伴って社会の諸領域において随時発生する。その特徴はデータ生産者の手を離れても、普遍性を持った認識産物として、real worldの再構成と認識の用に供し得る点にある。ここでは疾病現象と人口現象の側面から論じたが、同様の論議は他の領域のデータについても適用が可能である。蓄積性と発展性を持った客観的知識の体系はこのようなデータ環境の整備が前提になるべきであろう。

7. 文献

- 1) 倉科, 丹後, 大林: 肝硬変死亡の変遷に関する研究. 日衛誌 39, 819-830, 1984
- 2) 川口, 宮前, 丹後, 倉科: わが国における疾病構造の形成過程に関する研究(第7報) 自殺の経年的推移とそれに関与する諸要因の分析. 日衛誌 40, 373, 1985
- 3) 倉科, 大林, 丹後, 宮前: 同上(第4報). 特異な死亡傾向を示す世代の存在とそれに関与する諸要因の分析. 日衛誌 40, 358, 1985
- 4) 倉科, 灘岡, 神沼: 疾病構造解析システム-SAGE-. 医用電子と生体工学 23 (特別号), 1985
- 5) 倉科, 丹後, 秦: 日本人の疾病構造の形成過程に関する研究(第1報). 総合的展望. 日衛誌 39, 118, 1984
- 6) 丹後, 秦, 倉科: 同上(第3報). 人口動態統計に基づく出生コホートの問題点. 日衛誌 39, 120, 1984
- 7) Case RAM: Cohort analysis of mortality rates as an historical or narrative technique. Brit J. Prev Soc Med 10, 159-171, 1956