

厚生科学研究費補助金
(統計情報高度利用研究事業)

健康危機の早期検出を目指した厚生統計情報
等の自動解析システムの開発に関する研究

平成 1 1 年度研究報告書

主任研究者 山口 直人

厚生科学研究費補助金（統計情報高度利用研究事業）
総括研究報告書

健康危機の早期検出を目指した厚生統計情報等の自動解析システムの開発に関する研究

主任研究者 山口 直人 国立がんセンター研究所がん情報研究部長

研究要旨 疾病の地域集積を早期かつ的確に検出するために厚生統計情報等を自動解析するシステムの構築に関する研究を行った。本システムにより、これまでに解明されていなかった疾病リスクの存在を明らかにすることで、迅速な予防対策の実現につながると期待される。本年度は、疾病の地理的・時間的な異常集積を市区町村単位で分析するための地図表示システムの研究・開発を行った。また、解析手法としては、人口流入と小児白血病リスクに関する地理学的分析、出生性比の分析等を行い、地理学的な解析の有用性を示すことかてきた。自動解析システムとしては、上述の地図表示システムの他に、市区町村別に死亡構造を死因別死亡率の形で経年変化も含めて提示するシステムの開発を行った。本システムはインターネット上での公開が可能で、地域における疾病対策の基礎資料を提供するという点でも有用性が期待できる。

分担研究者

- | | | |
|-------------------|-----|------|
| 1. 国立がんセンター研究所 | 部長 | 山口直人 |
| 2. 東京都老人総合研究所 | 室長 | 水野正一 |
| 3. 鹿児島大学医学部 | 教授 | 秋葉登白 |
| 4. 広島大学原爆放射線医学研究所 | 教授 | 大籠 慈 |
| 5. 広島県立保健福祉短期大学 | 助教授 | 川崎裕美 |

A. 研究目的

本研究の目的は、疾病の異常な集積を市区町村単位のレベルで、早期に、かつ的確に自動検出するためのシステムを構築することである。人口動態調査、患者調査、地域がん登録などのデータは従来、経年変化や都道府県別比較などの広域的な集計を中心に解析されており、わかりの厚生行政に貴重な資料を提供している。しかし、近年、比較的小地域に悪性腫瘍の集積を疑う意見が出され、環境中の化学物質との関連が科学的根拠のないままに指摘されるなど、市区町村単位あるいはそれ以下のレベルでの疾病の集積を科学的な方法で把握する必要が急速に高まっている。本研究は、死亡、死産、罹患等のデータを定期的に解析し、小地域における真の地域集積を自動検出するシステムを開発することを旨とするが、これによって、市民に対する「健康危機」に早期に、かつ的確に対応することが可能になり、逆に不確かな情報で、存

在しない健康危機が無用の混乱を起こすことを回避できる。本研究での開発は、研究終了時点での実用化を目指すものであり、しかも過去に遡って解析を行い、現在の問題にも対応できるものである。

B. 研究方法

研究は、（１）既存の利用可能な疾病情報のデータベース構築、（２）自動解析システムの開発、（３）4kmメッシュを基準とした地理情報システム（GIS）の活用に関する検討、（４）医療施設の患者データベースを活用する可能性に関する検討の４つの分野に分けて行う。このうち、（１）データベース構築に関しては、人口動態死亡調査のデータベース構築を市区町村単位で行う。死亡数の解析には各市区町村の人口データが必要だが、人口データは国勢調査年のデータを基本とし、調査年の人口は性・年齢別に内挿処理を行って推定する。まず死亡データベースを完成させて、解析への利用可能性を検討し、確認した後に患者調査データ、人口動態死産データ等に拡張する。（２）自動解析システムの開発では、まず、様々なテーマについて、実際に地理集積性の解析を実施することで、自動解析システムに氷められるシステム要件を検討し、その結果を基に、市区町村単位で経年変化と地理分布の双方を同時に解析できるシステムを

開発する。さらに、特定の市区町村に注目して、その人口動態統計、特に、死亡について、データベースから参照可能なシステムを構築する。(3) GIS の活用に関しては、人口動態調査データ等を 4km メッシュデータとして集計することの実行可能性、実行してきた場合に市区町村レベルの解析との利点・問題点の比較等を理論レベルでの検討、シミュレーション実験による検討、一部小地域において実際に死亡データをメッシュデータに整理集計する試行を行って検討する。(4) 医療施設の患者データベースの活用については、その利用可能性に関する検討を中心に行う。計画段階での想定としては、第一段階の自動解析は人口動態調査、患者調査データ等を中心に行い、集積の可能性が半明した段階で、医療施設の患者データベース、健康保険レセプトデータ等に戻って、疾病の詳細に関してデータ収集を行うことが実行可能かどうか、問題点を検討する。また、一部小地域において、実際に死亡データから医療施設への患者データを照会、調査する試行を行って、問題点の洗い出しを行い、特に、患者のプライバシー保護の保証など、具体的な問題点を検討する。

C 研究結果

(1) データベース構築

1975年から1994年までの20年間の死因別死亡を、性・5歳年齢階級別・市区町村別に「全国死亡データベース」として整備した。また、人口についても、同時期の性・5歳年齢階級別・市区町村別人口をデータベース化した。人口は1975年調査年のデータを基に、その間の人口を内挿法によって推定した。特に、この期間に起こった市町村合併を自動的に考慮する設計を行った。データベース管理システムはリレーショナルデータベースを用い、死因分類等のマスターテーブルを用いて各種の分析かてきるように配慮した。今後、国際疾病分類が新しくなった1995年以降のデータも、ICD9とICD10の対応関係を考慮しつつ、データベースに追加する予定である。

(2) 地理学的自動解析

急速な人口流入が起りつつある地域で小児白血病が増

加するという仮説の検討を昨年度に引き続いて行った。1975年と1980年の人口を比較して、5年間に占める人口増加率を求め、1983-87年の小児白血病死亡率と人口増加率の関連をポアソン回帰分析によって解析した。人口規模が5千人以上5万人以下の小都市と5万人以上50万人以下の中都市とに分け、白血病死亡率も0-4歳、5-9歳、10-14歳の3群に分けて解析した結果、いずれの年齢階級でも人口増加率が高い市区町村の方が、白血病死亡率が高い傾向が認められた。この傾向は、中都市における10-14歳の死亡では統計的にも有意であった。

出生に関する分析としては、出生性比の経年変化に関する分析を行った。解析に際しては家族内の子どもの数や出生順位の影響を受けないように第一子のみを対象とし、全国の市区町村を人口規模で、小規模群(8千人未満)、中規模群(8千人以上2万人以下)、大規模群(2万人以上100万人未満)の3群に分けて分析を行った。その結果、中規模群及び大規模群では1975年から1995年までの20年間に、女児の出生割合が統計的に有意に上昇する傾向が認められた。環境要因の変化による影響も考えられ、今後、さらに詳細な検討を進める予定である。

(3) 解析システム

解析システムとしては、経年変化と地理分布を同時に解析する自動解析システムの開発、解析結果等から注目された特定の市区町村について情報を分析するシステムの開発、様々な疫学解析の場面で必要となるデータを全国死亡データベース等のデータベースから効率的に抽出するシステムの開発を行った。

自動解析システムに関しては、昨年度、肺がんに関して行った地理学的解析を基に、全国の任意の市区町村に関して、主な部位の癌死亡の地理分布を、経年変化も加味して分析するプロトタイプシステムの開発を行った。今年度は、地図上に市区町村の標準化死亡率を表わして、視覚的に地理分布を分析するシステムを構築し、肺がん、胃がん、肝がん、結腸がんへの関心を深めた。来年度は、統計的な解析結果もシステムに加えて自動解析システムとしての有用性を高める予定である。また、

出生や患者調査などのデータについても順次、データベース構築を進めて行く。

さらに、上述のような解析を行った結果、特定の市区町村における地理集積が疑われた場合、その市区町村の死亡状況を分析できるように、全国の任意の市区町村の死因別死亡を性・5歳年齢階級別に表示するシステム「Life Web」プロトタイプシステムを開発した。本システムはインターネット上で利用が可能であり、利用者からの指示に応じて、「全国死亡データベース」から必要なデータを自動抽出して、性・年齢階級別の一覧表表示を行うほか、全国との比較結果を標準化死亡比の形で表示する。また、結果をグラフ表示することも可能である。さらに、特定の市区町村における死因別死亡の経年変化を調整死亡率の形でグラフ表示する機能も開発した。今後、個人情報保護のための必要な機能を追加するなど、さらに検討を進める予定である。

地理疫学的分析に必要なデータを全国死亡データベースから自動抽出するシステムは、上述の Life Web で開発した諸機能を活用して、市区町村、性、年齢等を指定して、データを自動抽出する仕組みを開発する予定である。

D. 考察

本年度は、昨年度に引き続き、本研究の最終目的である自動解析システムの開発に向けた準備として、全国死亡データベースの構築、具体的な地理疫学的解析の実施、解析システムのプロトタイプの開発を行った。データベース構築に関しては全国死亡データベースの構築を来年度中に終える予定である。さらに来年度中には患者調査など、他のデータについてもデータベース構築を開始する。解析システムとしては、地理分布の自動解析システムについてプロトタイプシステムの開発を終えることかてき、地理分布を視覚的に分析できるようになった。この成果を受けて、視覚的な分析を裏付ける統計的解析を付け加えて、解析システムとして完成させる予定である。また、特定の市区町村の死亡構造を分析する Life Web もプロトタイプシステムが完成した。今後の予定として、視覚的な分析、統計的解析、そして特定の市区町村の分析の3者を統合して、より高度な解析システムを

目指す予定である。また、Life Web をインターネット上での利用を想定して開発して成功したことを受けて、他のシステムも最終形としてはインターネット上での公開も視野に入れた検討を進める予定である。最後に、個々の地理疫学解析のための効率的なデータ抽出に関しては、幅広い解析のニーズに対応して、解析用データセットを自動的に生成するインターフェースの構築が重要であることか本年度の研究でも明らかにされたので、今年度までの検討結果を基に、抽出システムとして持つべきシステム要件を整理して、来年度中に開発に着手したいと考えている。

E. 結論

本年度は、市区町村別の全国死亡データベースを基に、各種の地理疫学的解析を行い、貴重な成果を得た。また、出生性比の分析を通して、他の人口動態統計データの有用性を示すこともできた。自動解析システムとして、全国の市区町村における死亡リスクを視覚的に分析できる地図表示システムを開発した。さらに、市区町村別の死因別死亡を詳細に参照できるシステム「LifeWeb」のプロトタイプを開発した。

F. 研究発表

論文発表

- 1) Kaneko S, Yamaguchi N. Epidemiological analysis of site relationships of synchronous and metachronous multiple primary cancers in the National Cancer Center, Japan, 1962-1996. Japanese Journal of Clinical Oncology. 29:96-105, 1999.
- 2) Kinjo Y, Beral V, Akiba S, Key T, Mizuno S, Appleby P, Yamaguchi N, Watanabe S, Doll R. Possible protective effect of milk, meat and fish for cerebrovascular disease mortality in Japan. Journal of Epidemiology. 9 268-274, 1999.
- 3) Yamaguchi N, Kinjo Y, Akiba S, Watanabe S. Ultraviolet radiation and health: from hazard identification to effective prevention. Journal of Epidemiology. 9 S1-S4, 1999.

- 4) Araki K, Nagano T, Ueda M, Washio F, Watanabe S, Yamaguchi N, Ichihashi M. Incidence of skin cancer and precancerous lesions in Japanese -risk factors and prevention-. Journal of Epidemiology 9 S14-S21, 1999.
- 5) Takahashi K, Pan G, Feng Y, Ohtaki M, Watanabe S, Yamaguchi N. Regional correlation between estimated UVB levels and skin cancer mortality in Japan. Journal of Epidemiology. 9 S123-S128, 1999.
- 6) Akiba S, Shinkura R, Miyamoto K, Hillebrand G, Yamaguchi N, Ichihashi M. Influence of chronic UV exposure and lifestyle on facial skin photo-aging. Results from a pilot study. Journal of Epidemiology. 9.S136-S142, 1999
- 7) Yamaguchi N. Uncertainty in risk characterization of weak carcinogens. In A J Bailar, C Maltoni, J C Bailar III, F Belpoggi, J V Brazier, M Soffritti (eds.) Uncertainty in the risk assessment of environmental and occupational hazards. Annals of New York Academy of Sciences 895 338-347, 1999.
- 8) Yamaguchi N, Abe K. Gastric cancer. D Harber, G L Blackburn, V L W Go (eds.) Nutritional Oncology San Diego, Academic Press, pp.477-487, 1999

G 知的所有権の取得状況

特になし。

厚生科学研究費補助金（統計情報高度利用研究事業）
分担研究報告書

患者データベースの活用方法の検討

主任研究者 山口 直人 国立がんセンター研究所がん情報研究部長

研究要旨 疾病の異常な集積を市区町村レベルで、早期かつ的確に自動検出するためのシステムとして、ある市区町村において疾病の異常集積が疑われた場合に、その詳細な情報を参照できるように、市区町村別に死因別死亡の詳細な情報を提供するシステムを開発することを目的として研究を行った。各市区町村別に各死因の死亡状況を、性・5歳年齢階級別に表形式で表示するシステムを開発した。表にはさらに全国を標準とした場合の標準化死亡率（SMR）を表示して、容易に解釈できるように工夫した。さらに、表の内容をグラフ表示して視覚的にデータを吟味できる工夫も加えた。経年変化の表示には、1980年モデル人口に基づく年齢調整死亡率の経年変化を男女別に表示する仕組みを作った。経年変化も表とグラフの双方でデータ表示ができるようにした。本分担研究が構築した LifeWeb は、市区町村単位での死因別死亡をデータベースから容易に抽出して分析できる点が画期的であり、地理分析システムと連携して利用することによって、市区町村単位での死亡の異常集積を詳細に分析することを可能にするものである。また、市区町村レベルで疾病対策を企画立案したり評価したりする際には、市区町村レベルでの死因別死亡の動向を分析することが貴重な情報をもたらすが、本システムは地域保健医療計画の作成と評価においても大きな力を発揮すると期待できる。

A. 研究目的

疾病の異常な集積を市区町村レベルで、早期かつ的確に自動検出するためのシステムを構築することが本研究の目的である。そこで、本分担研究では、ある市区町村において疾病の異常集積が疑われた場合に、その詳細な情報を参照できるように、市区町村別に死因別死亡の詳細な情報を提供するシステムを開発することを目的として研究を行った。本システムを地理集積性の自動検出システムと連携して利用することによって、死亡の異常集積を的確に発見して早期に対策をたてる体制を実現する。

は、大瀧班員と川崎班員が経勢調査データに基づいて内挿処理で推定した人口を用いた。これらの死因別死亡データと人口データを、市区町村名等の各種マスタとともに関係型データベースとして整備した。データベース管理システムは、マイクロソフトアクセスを用いた。さらに、完成した関係型データベースを Web 上での表示に対応させるサーバ、グラフ作成用サーバを導入して、WWW フラウザで市区町村別のデータを関係型データベースから抽出して自動的に集計し、Web 上で表示するシステムを開発した。

B 研究方法

死因別死亡データは、1975年から1994年までの人口動態調査調査死亡票を暦年別に、市区町村別・性別・5歳年齢階級別に集計した中間集計ファイルを SAS システムを用いて作成した。また、1975年から1994年までの市区町村別・性別・5歳年齢階級別の人口

C. 研究結果

各市区町村別に各死因の死亡状況を、性・5歳年齢階級別に表形式で表示するシステムを開発した。市区町村別の死亡数を暦年別に作表すると、死亡数、人口ともに少数になってしまうことから、5年分を集計して作表するようにした。表にはさらに全国を標準とした場合の標

準化死亡比 (SMR) を表示して、容易に解釈できるように工夫した。さらに、表の内容をグラフ表示して視覚的にデータを吟味できる工夫も加えた。表示に際しては個人情報保護の観点から性・5歳階級別の総死亡数から5人未満の場合は個人の死因が特定できる恐れを考慮して、死亡数、死亡率を表示しないようにした。

経年変化の表示には、1980年モデル人口に基づく年齢調整死亡率の経年変化を男女別に表示する仕組みを作った。経年変化も表とグラフの双方でデータ表示ができるようにした。

システムの表示画面と遷移形式を図1に示した。

D. 考察

疾病の地理集積性の分析は、異常集積を早期に発見して的確な対策を講じる上で極めて重要である。従来の分析では都道府県レベルでの解析が主体であって、市区町村レベルでの分析は困難であった。本分担研究が構築したLifeWebは、市区町村単位での死因別死亡をデータベースから容易に抽出して分析できる点が画期的であり、地理分析システムと連携して利用することによって、市区町村単位での死亡の異常集積を詳細に分析することを可能にするものである。また、市区町村レベルで疾病対策を企画立案したり評価したりする際には、市区町村レベルでの死因別死亡の動向を分析することが貴重な情報をもたらすか、本システムは地域保健医療計画の作成と評価においても大きな力を発揮すると期待できる。

本システムはWebを基本として開発されており、インターネット上に比較的容易に公開することが可能である。来年度中には試験的にインターネット公開して、利用者の意見をフィードバックするなどして、よりよいシステムを目指す予定である。

E. 結論

市区町村別の死因別死亡の状況を詳細に分析するためのシステムとして、LifeWebを開発した。本システムでは、年齢階級別の死亡を表、グラフの形式で参照できるほか、1975年から1994年までの死亡の経年変化

を年齢調整死亡率の形で分析できる機能も付加した。本システムは死亡の異常集積の分析において大きな力を発揮すると期待できるほか、地域における保健医療計画の策定と評価に際しても活用が期待できるものである。

F. 研究発表

- 1) Kaneko S, Yamaguchi N Epidemiological analysis of site relationships of synchronous and metachronous multiple primary cancers in the National Cancer Center, Japan, 1962-1996. Japanese Journal of Clinical Oncology. 29 96-105, 1999.
- 2) Kinjo Y, Beral V, Akiba S, Key T, Mizuno S, Appleby P, Yamaguchi N, Watanabe S, Doll R. Possible protective effect of milk, meat and fish for cerebrovascular disease mortality in Japan. Journal of Epidemiology 9 268-274, 1999.
- 3) Yamaguchi N, Kinjo Y, Akiba S, Watanabe S Ultraviolet radiation and health from hazard identification to effective prevention. Journal of Epidemiology 9 S1-S4, 1999.
- 4) Araki K, Nagano T, Ueda M, Washio F, Watanabe S, Yamaguchi N, Ichihashi M Incidence of skin cancer and precancerous lesions in Japanese -risk factors and prevention- Journal of Epidemiology 9 S14-S21, 1999
- 5) Takahashi K, Pan G, Feng Y, Ohtaki M, Watanabe S, Yamaguchi N. Regional correlation between estimated UVB levels and skin cancer mortality in Japan. Journal of Epidemiology 9 S123-S128, 1999.
- 6) Akiba S, Shinkura R, Miyamoto K, Hillebrand G, Yamaguchi N, Ichihashi M Influence of chronic UV exposure and lifestyle on facial skin photo-aging Results from a pilot study Journal of Epidemiology. 9 S136-S142, 1999.
- 7) Yamaguchi N. Uncertainty in risk characterization of weak carcinogens In A J

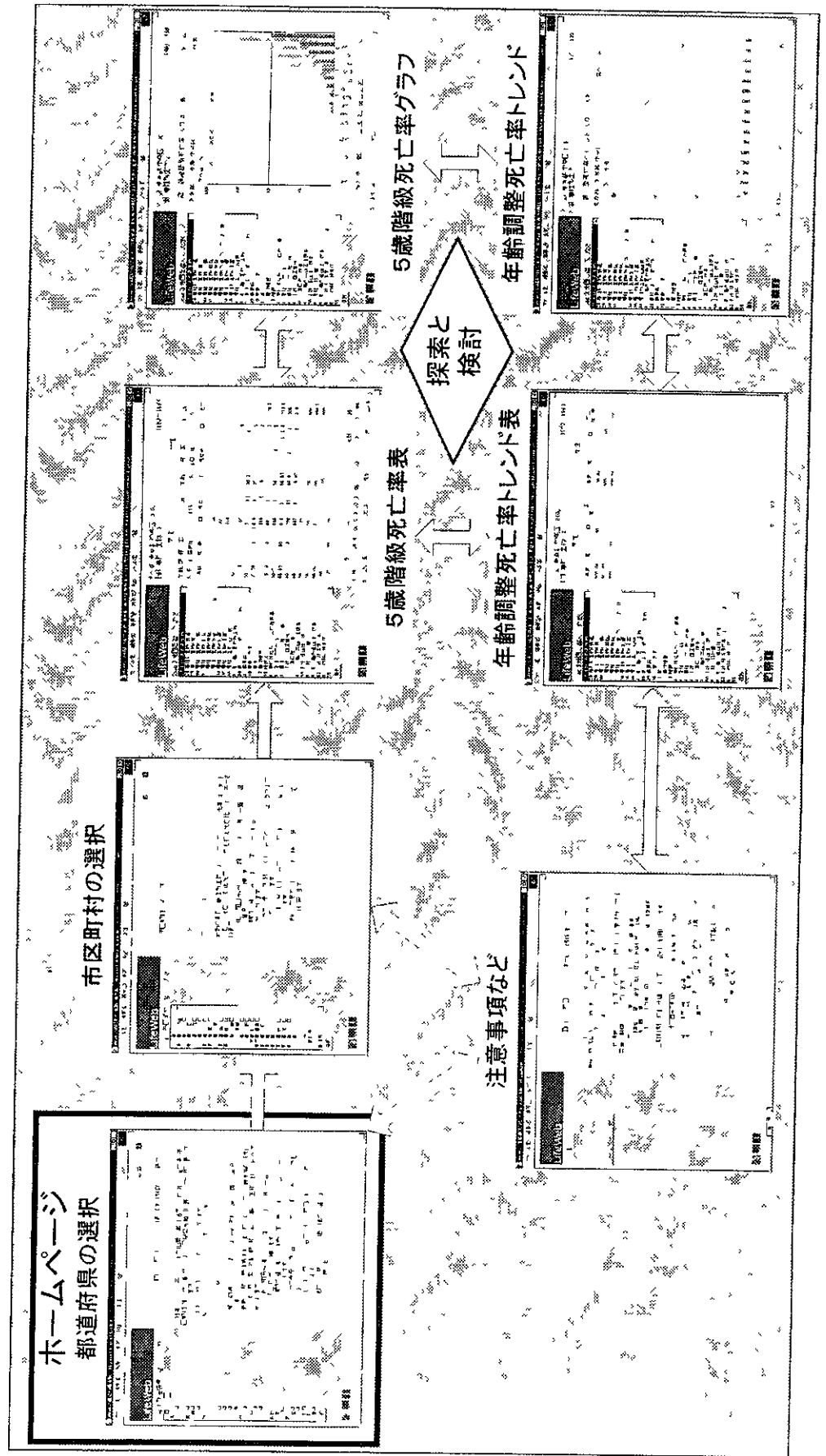
Bailar, C Maltoni, J C Bailar III, F Belpoggi, J V Brazier, M Soffritti (eds.) Uncertainty in the risk assessment of environmental and occupational hazards Annals of New York Academy of Sciences. 895.338-347, 1999.

8) Yamaguchi N, Abe K. Gastric cancer. D Harber, G

L Blackburn, V L W Go (eds.) Nutritional Oncology San Diego, Academic Press, pp.477-487, 1999

G. 知的所有権の取得状況
特になし。

図1. LifeWebにおける画面遷移の概要



出生児性比の経年変動に関する統計解析

分担研究者 大瀧 慈 広島大学原爆放射能医学研究所・教授

研究要旨・近年、ダイオキシンなどの環境ホルモンによるヒトへの健康影響や生態系の破壊に関して、懸念されている。その一つに出生児における性比の下降傾向が挙げられる。本研究では、1975年～1997年の期間に日本国内での全ての出生データを基に、市区町村別に男児と女児の人数を集計し、男女合計を固定した場合女児人口数が日本全国の市区町村で共通な比率母数を持つ二項分布に従っているか否か、女児割合の経年変化が存在するか否か、という問題について統計的方法を適用し検討を行った。

A. 研究の目的

本研究の目的は、「最近の日本における出生児性比に経年的変動や地理的変動が存在するか否か？」という問題について最新の統計的手法を用いて解析し、検討することである。

4) 1975年度の各市区町村の総人口数を用いて、ほぼ等分になるように三段階の人口規模にクラス分けした（参照 Table 2）。他の年度についても同じクラスを使用した。

B. 研究方法と結果

B1 データと前処理

解析の対象となったデータは厚生省から目的外使用を許可され入手した1974年から1997年までの約20年の間に生まれた出生児データである。このデータを直接使用せずに次のような前処理を行った（参照 Table 1）。

1) 1975-79年、・・・、1995-97年のデータをそれぞれ1975年度、・・・、1995年度のデータとしてマージした。

2) 出生順位によるバラツキを排除するため第一子のみを抽出し、各市区町村ごとに女児数、男女数として集計した。

3) 各年度ごとに国勢調査による各市区町村の総人口数のデータを結合させた。

B2 全国における出生児性比の均一性に関する検証

Table 1. 解析の対象となったデータ

CityId	Female	Baby	Total	Class
34212	916	1957	113932	3
34302	452	927	50639	3
34304	336	675	30047	3
34307	139	265	24952	3
34309	60	112	12417	2
34310	67	128	14129	2
34311	63	138	16263	2
34312	33	58	8362	2
34313	7	16	3212	1
34314	8	17	3032	1

Class 1=[1,8千),2=[8千,2万), 3=[2万,100万)

(1995年度データより一部抜粋)

まずはじめに、全国の市区町村において女兒人口数が独立同一の二項分布に従うかどうかを検討した（参照 図 1）。

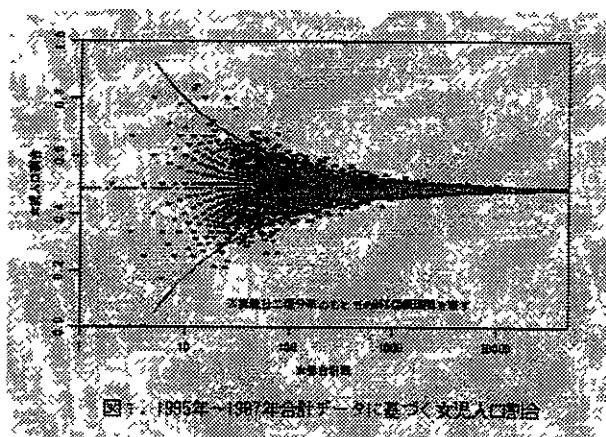


Table 2. 年度および人口規模別における市区町村数

Year	[1,8千)	[8千,2万)	[2万,100万)
1975	1181	1179	1009
1980	1181	1178	1006
1985	1181	1176	998
1990	1179	1172	986
1995	1177	1161	971

以下、簡単のため出生児性比の代わりに女兒人口割合（女兒人口数/男女合計数）を扱を扱う。市区町村ごとに「母集団の女兒人口割合が異なる」という状況を想定し、これをベータ二項分布として表現した。ここで、二項分布における試行回数、成功数および成功確率は、それぞれ、ある市区町村における男女合計数、女兒人口数および女兒人口割合に対応している。しかしながら、二項分布における試行回数は固定されたものであり、男女合計数の場合と違い、確率変数ではない。

このモデルのもとで帰無仮説「女兒人口割合は全国で均一」を年度別人口規模別の各層ごとに検定した。その結果、ほとんどの層において帰無仮説は棄却されなかった（参照

Table 3, 4）。

Table 3. ベータ分布の標準偏差の最尤推定量

Year	[1,8千)	[8千,2万)	[2万,100万)
1975	0.00153	0.00808	0.00182
1980	0.00417	0.00000	0.00211
1985	0.00768	0.00000	0.00099
1990	0.00589	0.00840	0.00129
1995	0.01355	0.00000	0.00018

Table 4. 検定統計量の p 値

Year	[1,8千)	[8千,2万)	[2万,100万)
1975	0.4878	0.0105*	0.0774
1980	0.4190	0.5000	0.0572
1985	0.2866	0.5000	0.3921
1990	0.3792	0.0283*	0.3037
1995	0.1752	0.5000	0.4977

* $p < 0.05$ ただし、 p 値は各層における有意水準を表す。

B3 人口規模別市区町村別性比の経年変化

この節では、我々の仮説である「女兒人口割合の経年変化は人口規模により異なる」をロジスティックモデルを用いて検証する。前節の結果から、全国の女兒人口数 Y_{jk} は男女合計数 N_{jk} 、女兒人口割合 P_{jk} を用いて二項分布 $B(N_{jk}, P_{jk})$ （ただし、 $j = 1, \dots, 5$ は、それぞれ、1975 年度、・・・、1995 年度を表し、 k は人口規模のクラスを表す）に従うと考えることができる（参照 Table 5）。

Table 5. 女兒人口割合の観測値

Year	[1,8千)	[8千,2万)	[2万,100万)
1975	0.485286	0.485442	0.485215
1980	0.486439	0.487288	0.485769
1985	0.484407	0.486119	0.485684
1990	0.485580	0.487292	0.485708
1995	0.488249	0.487627	0.486051

経年変化として2次の多項式を仮定した次のロジスティックモデルを考える。

$$\log \frac{p_{jk}}{1-p_{jk}} = \beta_0 + \sum_{l=1}^9 \beta_l z_l,$$

ただし、 β_0 および β_l 's は未知回帰係数パラメータ、 z_1, z_2, z_3 は k 番目の人口規模クラスを表すダミー変数、*i.e.*,

$$z_1 = \begin{cases} 1, & k=1 \\ 0, & \text{else} \end{cases},$$

$$z_2 = \begin{cases} 1, & k=2 \\ 0, & \text{else} \end{cases},$$

ここで、 $z_3 = 1 - z_1 - z_2$ が成り立つことから z_3 それ自身は解析に用いていないことに注意する。また、 z_l ($l=4, \dots, 9$) は各クラスごとの経年変化に関連した次の変数である。

$$z_4 = z_1 t_j, \quad z_5 = z_2 t_j, \quad z_6 = z_3 t_j,$$

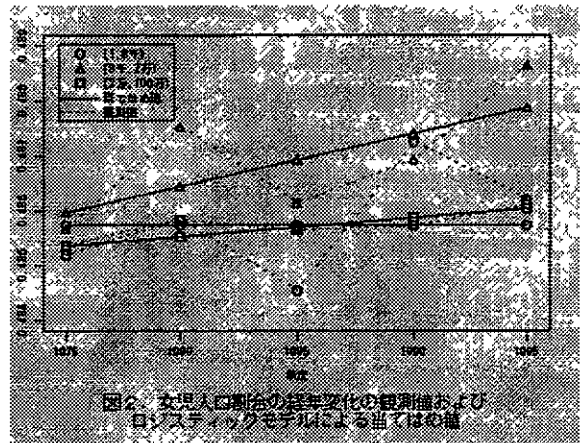
$$z_7 = z_1 t_j^2, \quad z_8 = z_2 t_j^2, \quad z_9 = z_3 t_j^2,$$

$$t_j = \{-10, j=1; -5, j=2; 0, j=3, 5, j=4, 10, j=5\}.$$

上記モデルの基で AIC による変数選択の結果、人口規模により異なる上昇傾向が示唆された (参照 Table 6, 図 2)。

Table 6. AIC によって選択されたロジスティックモデル estimates 95% confidence intervals

	estimates	95% confidence intervals	
β_0	-5.722×10^{-4}	$(-5.816 \times 10^{-4},$	$-5.628 \times 10^{-4})$
β_1	1.090×10^{-4}	$(-4.409 \times 10^{-3},$	$4.627 \times 10^{-3})$
β_2	4.965×10^{-3}	$(2.159 \times 10^{-3},$	$7.771 \times 10^{-3})$
β_5	3.830×10^{-4}	$(3.230 \times 10^{-6},$	$7.628 \times 10^{-4})$
β_6	1.402×10^{-4}	$(5.347 \times 10^{-6},$	$2.751 \times 10^{-4})$



さらに、パラメトリックなモデルを使うことでのバイアスを考慮して人口規模別にノンパラメトリックな傾向性の検定 (参照 Bartholomew (1959), Kulatunga et al (1996)) を行った。その結果、中規模、大規模な市区町村において女児人口割合の上昇傾向が認められた (参照 Table 7)。

Table 7. 検定統計量の値とその p 値

Year	[1,8千]	[8千,2万]	[2万,100万]
T_{II}	1.0375	4.6054	4.1492
p-value	0.3893	0.0649	0.0803

* $p < 0.05$

C. まとめ

一般に、胎児の男児は女児に比べ生存率が低いことが知られている。しかしながら、昨今の医療技術の進歩により格段にこの状況は改善されているので、出生児性比は増大してきても不思議ではない。それにもかかわらず、本研究の結果、出生児性比が減少していることが、示唆された。

D. 研究結果の公表

未公表 (平成 12 年度日本計量生物・応用統計学会合同年会 (5 月 12 日) にて発表予定)

統計情報高度利用のための自動データベース整備システムの開発に関する研究

分担研究者 川崎裕美 広島県立保健福祉短期大学・助教授

研究要旨 健康問題に関わる環境因子を探索するためのコンピュータシステム構築 整備の一環として、多量の疾病情報データを統計処理 解析目的に適した構造で自動的にデータベース化するシステムを、長期にわたる市区町村別のデータである地域がん登録を利用して検討した。

A. 研究目的

長期にわたる地域がん登録データの統計処理に適した構造のデータベース化、特に様々な解析目的に適したデータベースを作成する方法を検討している。これまでの研究から、研究者の解析目的によって対象データが異なること、たとえば短期間でも区単位のごく小地域のデータがほしい、政令市全体でよいから長期間のデータがほしいといったようなことがデータベースの整備に求められることが想定された。目的に合わせた市区町村の状態データベースが取り出せるような工夫の必要性が明らかとなった。

そこで、これまでに整備した市区町村別年齢別の5歳階級の累積死亡数および人口データベースから県単位や政令市単位で年次別死亡数および人口を求める汎用性に富んだコンピュースプログラムの開発を試みた。

B. 研究方法および結果

集計単位の任意変更を可能にするための手順

これまでに整備した市区町村別の20年間の性別5歳階級の累積死亡数および人口データベースの集計市区町村を任意に変更するためには、年次別データに変更、任意の市町村の集計、市町村別データに変更という3つの段階を必要とした。

1 市区町村別データを年次別データに変換するプログラムの作成

年度別地域がん登録の死亡票から市区町村毎に20年間の死亡数が5歳階級別に記録された表1のようなデータベースを作成することがこれまでの研究から可能になった。これをもとにして県単位・政令市単位で年度別に市区町村死亡数を加算するために、市区町村別年次死亡数をソートし、年次別のデータファイルを作成する。推定人口を含む20年間の人口データベースについても同様の処理が行えるようにした。

2. 集計単位を任意に変更するためのプログラムの作成

県単位の死亡数の算出の場合は目的のがんに関するデータベースから市区町村コードの各県のコードにあたる部分に従って加算する。また、政令市の場合には、政令市に所属する区を参照ファイルに従って選び出し、加算していく。参照ファイルを作成することによって研究者が合算したい市区町村コードを記入することで、任意に市区町村のデータを集計する事が可能となった。

3. 年次別データを市区町村別データに変換するプログラムの作成

年次別になっている20個の中間ファイル、市区町村別5歳階級の20年間データを結合し、県あるいは政令市コードによってソートする。この作業によって同じ地区の20年間の死亡数を順に並べることができる（表2）。

4. バッチシステムによる処理システムの自動化

これまでの1から3の処理をすべて自動化し、一連のパッケージによるバッチ処理で実現可能とした。

C. 考察

1998年次点での全国の市町村別のデータベースを、研究者が望む条件に整備するという課題のうちの特定の市区町村を取り出すことは可能となった。参照ファイルを作成することで、今後政令市の区の再編が行われても簡単に政令市としての集計を変更することができるようになった。

今後の課題は、1998年次点での市区町村の状態を示される過去のデータをその年次での市区町村の状態で活用したい場合、つまり解析期間の自由選択に対応できることである。今後も行われる市区町村の合併および分離への対応も行っていく必要がある。

悪性新生物による死亡だけでなく、全死因の死亡数について、同様の整備を進めているところであるが、悪性新生物死亡と同様ICD-10へのコード変換を進め1995年以降のデータ利用を可能にすることが必要である。

D. 研究発表

学会発表

川崎裕美，佐藤健一，中山晃志，山口直人，大瀧慈 ノンパラメトリック回帰分析による肺がん死亡率の経年変動に関する地理的特性の検討 日本計量生物学会 応用統計学会，東京，1999

表1 気管・気管支・肺の悪性新生物年齢階級別市町村年次別死亡数

調査年	市町村 コード	年 齢 階 級										
		0-4	5-9	10-14	15-19	20-24	70-74	75-79	80-84	85-89	90-94	95-
1975	1101	0	0	0	0		11	6	0	0	0	0
1976	1101	0	0	0	0		15	5	5	0	0	0
1977	1101	0	0	0	0		10	9	5	8	2	8
1978	1101	0	0	0	0		9	4	13	9	0	0
1979	1101	0	0	0	0		6	6	16	15	0	0
1986	1102	0	0	0	0		10	6	23	4	0	0
1987	1102	0	0	0	0		18	21	19	5	0	0
1988	1102	0	0	0	0		16	10	15	9	0	0
1989	1102	0	0	0	0		11	8	14	10	1	0
1990	1102	0	0	0	0		21	6	3	15	2	0
1991	1102	0	0	0	0		19	11	12	9	5	0
1992	1102	0	0	0	0		20	15	19	7	0	0
1993	1102	0	0	0	0		18	9	21	19	0	0
1994	1102	0	0	0	0		22	5	11	8	0	0

表 2 気管・気管支・肺の悪性新生物年齢階級別県別年次死亡数

調査年	県 コード	年 齢 階 級										
		0-4	5-9	10-14	15-19	20-24	70-74	75-79	80-84	85-89	90-94	95-
1975	1	0	0	0	0		102	65	28	5	0	0
1976	1	0	0	0	0		122	79	31	3	3	0
1977	1	0	0	0	0		128	85	44	13	11	9
1978	1	0	0	0	0		176	104	58	13	3	0
1979	1	0	0	0	0		168	117	61	18	2	0
1986	1	0	0	0	0		70	65	55	15	1	0
1987	1	0	0	0	0		66	72	51	18	1	0
1988	1	0	0	0	0		67	77	49	14	2	1
1989	1	0	0	0	0		76	86	56	16	7	0
1990	1	0	0	0	0		74	93	69	16	3	0
1991	1	0	0	0	0		81	85	58	29	11	1
1992	1	0	0	0	0		98	99	75	26	3	0
1993	1	0	0	0	0		90	97	71	24	6	0
1994	1	0	0	0	0		101	123	76	36	7	0

都市部からの急激な人口移入が起きている地域で小児白血病リスクが増加していないか？
鹿児島大学医学部
秋葉澄伯 孫全富

研究要旨

本研究では、昨年度に引き続き、都市部からの急激な人口移入が起きている地域で小児白血病リスクが増加していないかを検討した。昨年度は 1983-87 年までの死亡で検討を行ったが、今年度は観察期間を 5 年間延長して 1992 年までとした。解析結果からは昨年度までと同様に、人口規模（1980 年時点）が千以上 5 万未満の市町村の人口急増地域で、2-3 歳にかけて多く発生する急性リンパ性白血病が増加している可能性が示唆された。今後、観察期間を延長するなどして更に詳細に検討したいと考えている。

A 研究目的

近年、イギリスなど各国の都市部からの急激な人口移入が起きている地域で小児白血病リスクが増加していると報告されている。このような現象が常に認められるわけ出ないことも報告されており、わが国では、同様の現象が起きているのかまた、起きていたのかは興味深い問題である。

この現象の詳細なメカニズムについても議論のあるところである。未知の白血病関連ウイルス（または、その他の病原体）に感染した都市部住民が農林漁村へ流入して住民（特に小児）へ感染を生じ、その結果、生に白血病を生ずるのではないかとの仮説が有力である。本研究は、わが国でも、都市部からの急激な人口移入が起きている地域で小児白血病リスクが増加していないかを検討しようとするものである。近年、小児白血病の致命率は急速に低下しており（参考までに図 1 に乳児白血病の死亡率の経年変化を示した）、このような検討には白血病罹患データが望ましいが、わが国ではごく限られた一部の地域を除いてデータが得られていないので、ここでは死亡データを用いた検討を行った。

B 研究方法

人口データ、死亡データともに広島大学原爆放射能医学研究所大瀧教授から提供を受けたものを用いた。昨年度は 1983 年から 87 年の死亡データを用いて解析を行ったが、今年度はデータの整理が更に進んだため 1992 年までのデータを用いることが可能となった。従って、本研究では、1975 年と 1980 年を比へた人口比（ $=1980 \text{ 年人口} - 1975 \text{ 年人口}$ ）と 1983-92 年の小児白血病死亡率との関連をポアソン回帰分析で検討した。

C 研究結果

人口流入は中小都市で起きやすいことが予想され、実際に 1980 年と 1975 年の人口比（1980-1975）と人口規模の関連を見てもその予想が妥当であることが示されたので、対象を人口千人以上 50 万未満とし、さらに、これを人口 5 万未満と 5 万以上に分けて検討した。

表 1 から明らかなように人口規模（1980 年時点）が千以上 5 万未満で 1975 年から 1980 年の間に人口が増えている市町村では 0-4 歳の白血病死亡が増加している（統計学的には有意でない）か、5 歳以上の白血病死亡ではそのような傾向は認められない、これは 2-3 歳にかけて多く発生する急性リンパ性白血病が人口急増地域で増加している可能性を示唆するものである。一方、人口規模（1980 年時点）が 5 万以上 50 万未満で 1975 年から 1980 年の間に人口が増えている市町村では 10-14 歳の白血病死亡率が減少している。

表 2 ではどの年齢層の人口増加が 0-4 歳の白血病死亡率と関連しているかを検討したが、15-64 歳の年齢層の人口増加率が 0-4 歳の白血病死亡率と関連している可能性が示唆された。

D 考察

小児白血病のリスク要因としては様々な要因が疑われている。主なものは以下の要因である。

母親の妊娠時

- X 線検査等による放射線への曝露、
- 農薬等の化学物質への曝露、
- ウイルス感染、
- 薬の服用、
- 喫煙、
- 電磁場への曝露、

出世後

- 放射線曝露、
- 農薬等の化学物質への曝露、
- ウイルス感染、
- 受動喫煙、
- 電磁場への曝露

遺伝的要因

- ダウン症等の疾患

母親の妊娠時の特徴

- 出産年齢
- 生下時体重、

疑われている要因は上記のこく多岐に亘っているが、白血病への関与がはっきりしているものは少なく、出世後の高線量放射線被曝などごく限られた要因に過ぎない。近年、特に疑われているが population mixing によるウイルス感染である。イギリスなどで都市部からの急激な人口移入が起きている地域で小児白血病リスクが増加の見られることが少なくないのは都市部で未知の白血病関連ウイルスに感染した成人が非都市部に流入すると、その地域に居住していて白血病関連ウイルスに感染していない子供たちか感染を受けその後、白血病を起こすのではないかとの仮説である。本研究では、人口規模（1980年時点）が千以上5万未満の市町村の人口急増地域で、2-3歳にかけて多く発生する急性リンパ性白血病が増加している可能性が示唆された。この現象が白血病罹患でも見られる現象かどうかは興味深いか、既に述べたようにわか国では一部の地域を除いて、そのようなデータの入手は困難である。今後、観察期間を延長するなどして更に詳細に検討したいと考えている。

E 結論

本研究では、人口規模（1980年時点）が千以上5万未満の市町村の人口急増地域で、2-3歳にかけて多く発生する急性リンパ性白血病が増加している可能性が示唆された。今後、観察期間を延長するなどして更に詳細に検討したいと考えている。

F 研究発表

- 1 論文発表 なし
- 2 学会発表 なし

表1 人口増加率と白血病死亡率

年齢	全人口	全人口の 増加率 1980/1975	小児白血病(1983-92)			傾向性のP値
			死亡数	死亡率 (10万当)	RR 95%CI	
男女 0-4才	千から5万	0.9-	86	1.311	—	0.271
		1.0-	148	1.4211	0.8-1.4	
		1.1-	32	1.3310	0.7-1.5	
		1.2-	33	1.8214	0.9-2.1	
	5万から50万	0.9-	83	1.331	—	0.991
		1.0-	339	1.6212	1.0-1.6	
1.1-		116	1.3710	0.8-1.4		

男女 5-9 才 千から 5 万	1 2-	34	1 180 9	0 6-1 3	0 893
	0 9-	134	1 731	—	
	1 0-	214	1 751 0	0 8-1 3	
	1 1-	44	1 510 9	0 6-1 2	
	1 2-	42	1 801 0	0 7-1 5	

表2 人口増加率と小児白血病死亡率
男女 0-4 才 全人口千から五万

	小児白血病				傾向性の P 値
	死亡数	RR	95%CI		
0-14 歳人口増加率					0 533
	-0 9	162	1	-	
	1 0-	75	0 9	0 7-1 2	
	1 1-	25	0 9	0 6-1 3	
	1 2-	37	1 2	0 8-1 7	
15-64 才人口増加率					0 236
	-0 9	109	1	-	
	1 0-	134	1 1	0 8-1 4	
	1 1-	25	0 8	0 5-1 3	
	1 2-	31	1 5	1 0-2 3	