

環境因子の変化に伴う疾病構造変化モニタリングと中長期環境モニタリング及び  
曝露調査結果を用いた環境がヒトへ与える影響の解析を行う  
病院コホートを利用したデータマイニングシステムの研究

国立研究開発法人国立国際医療研究センター  
企画経営部研究医療課 情報管理室

星野隆之

平成24～28年度合計予算額 93,266千円  
(うち、平成28年度当初予算額 14,265千円)

[要旨] 温暖化をはじめとし、地球規模での環境変化による健康影響を調査するための基盤づくりに、各国では環境因子のモニタリングがさかんに実施され、いままで環境因子の影響が指摘されてこなかった、あるいは経験的に指摘されていても、エビデンスが得られていなかった疾病に関して、環境因子と疾病構造に関する研究が萌芽的に始まっている。我が国においても、全国を網羅し、一般公開されている環境ビッグデータが存在し、詳細な気象データやPM2.5を代表とする新たな環境因子が精密にモニタリングされている。そして徐々に新たな環境因子が疾病構造に影響を及ぼしていることが知られつつある。これら、従来研究の対象よりはるかに微細な環境因子の変動が、いかにわれわれの疾病構造に大きな影響をあたえているかを解析し、新たな知見を得るためには、いままで以上により大規模な症例群から、より詳細な臨床データを集めた、大きなデータベースが不可欠である。しかし、現在環境モニタリング結果と診療情報システムデータとを関連づけたデータ解析を行った事例は近年でもいまだに少数である。これは、診療データの記録目的が診療のためであり研究目的ではないことからデータに2次利用を行うためには様々な前処理が必要で、適切な解析システムが存在しなかったことが障壁となっている。

このため本研究では、情報科学と医学の集学的な取り組みにおいて、環境データと診療情報とを関連付けたデータ解析を行い、今後複数の協力施設と連携してデータ解析を行うことを念頭に、環境因子との関連性が強い疾病候補を絞り込み、継続的な環境因子の変動と疾病構造の変化のモニタリングを可能にするインフラとして、病院情報システムのデータと、環境省が保有する環境データ（環境省大気汚染物質広域監視システムのデータ等）の統合・解析を可能にする「共通システム」を構築した。

[キーワード] 環境因子、疾病構造解析、医療情報、電子カルテ、大気汚染、気象、ビッグデータ、解析システム、環境モニタリング

## 1. はじめに

本研究の目的は、複数の協力施設と連携してデータ解析を行うために、環境因子との関連性が強い疾病候補を絞り込むことが可能になるよう、継続的な環境因子の変動と疾病構造の変化のモニタリングを可能にするインフラとして、病院情報システムのデータと、環境省が保有する環境データ（環境省大気汚染物質広域監視システムのデータ等）の統合及び解析を実現する「共通システム」を構築することである。今年度は研究の最終年度であるため、本報告では本研究により得られた成果としてとくに下記3点について重点的に述べる。

- (1) 本研究を通じて構築した「共通システム」の全貌を解説する。
- (2) 本研究班が行った具体的な応用解析例について紹介する。
- (3) 今後複数の協力施設と連携してデータ解析を行うために「共通システム」を無償配布する目的で行う、学術的に公表・導入可能な施設への呼びかけについて説明する。

## 2. 研究目的

本研究では、情報科学と医学の集学的な取り組みにおいて、環境データと診療情報とを関連付けたデータ解析を行い、今後複数の協力施設と連携してデータ解析を行うことを念頭に、情報科学的な手法を駆使して、効率よくかつ個人情報に配慮して安全に、低コストで迅速な突合解析システムを構築するため、具体的な解析を通じて、環境因子との関連性が強い疾病候補を絞り込み、継続的な環境因子の変動と疾病構造の変化のモニタリングを可能にするインフラを構築することを目的としている。この実例として、病院情報システムのデータと、環境省が保有する環境データ（環境省大気汚染物質広域監視システムのデータ等）の統合・解析を可能にする一連のシステムや解析過程を統合、一部可能な部分を自動化した「共通システム」を構築する。構築した「共通システム」は、システム本体の対価をもとめず、本研究の目的である、今後複数の協力施設と連携してのデータ解析を実現するためのインフラ整備として、学術誌などを用いて一般に公開し、導入希望の医療施設等をつのり、希望施設があればシステムの提供を目指していく。

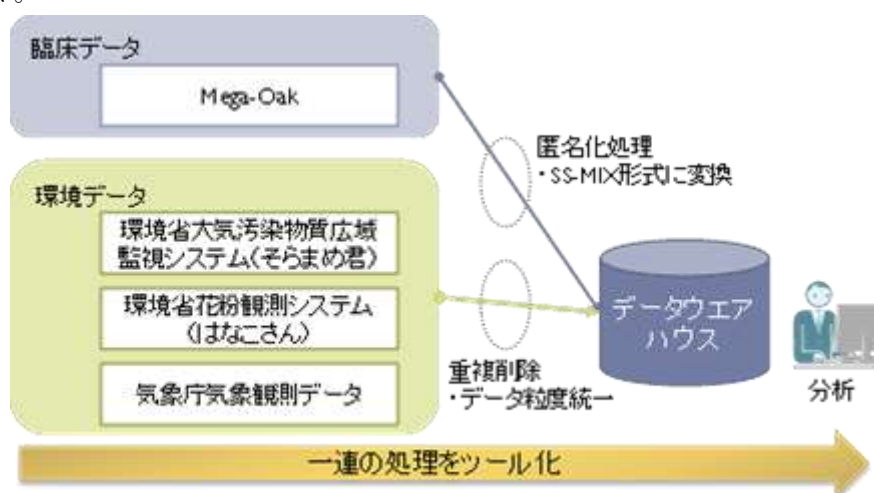


図 1 臨床データ及び環境データの取り込み・蓄積のイメージ

### 3. 研究方法

#### (1) システム設計方法

臨床データについては、病院情報システム内のデータを抽出、匿名化したうえで、標準化ストレージであるSS-MIXへとデータを取り込む。環境データについても、データを取り込んでSS-MIX側のネットワークに統合する。データ取り込みおよび蓄積のフローについては、2012年度に仕組み化した下図の成果を活用する。

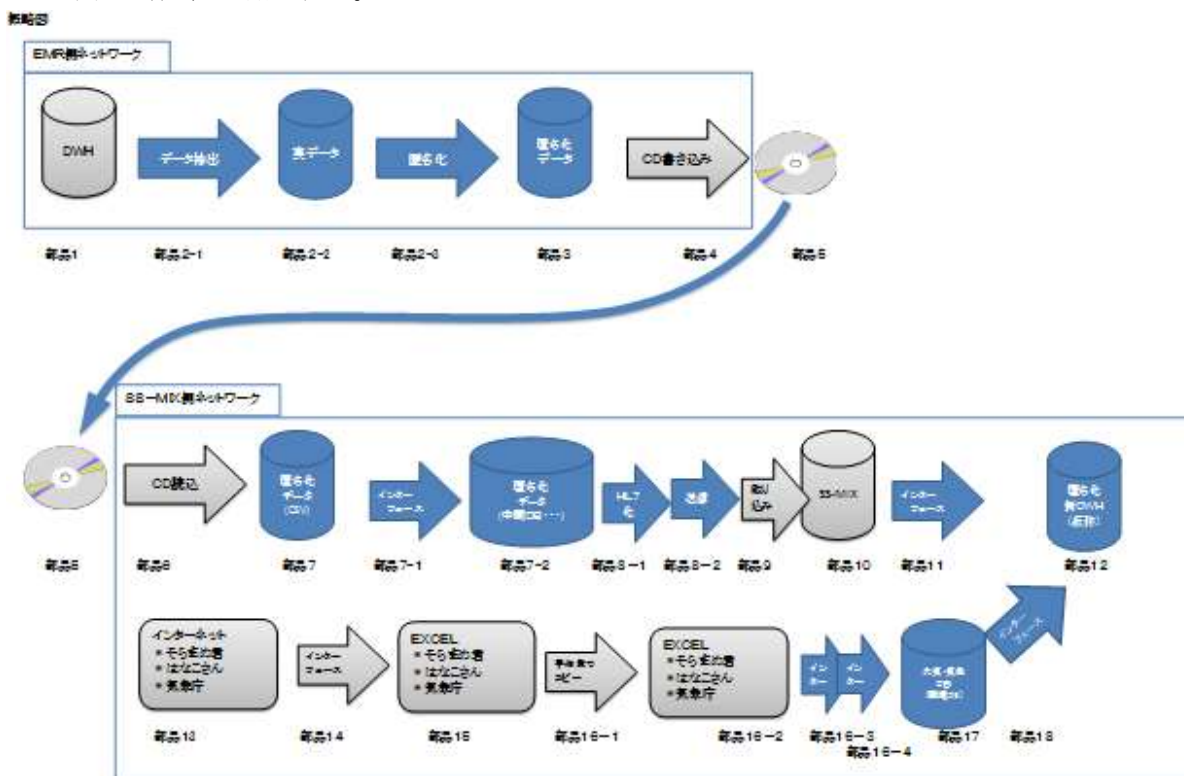


図 2 データ抽出・取り込みフロー（2012年度報告書より）

このフローでは、病院システムから抽出した実データは、匿名化処理を施された後CDに書き込まれる。次に、CDからデータを読み込んでHL7の形式に変換し、SS-MIXデータベースに取り込む。環境データはインターネット上からデータをエクセル形式で取り込み、重複データの削除、時間単位データから日単位データへの集約などの処理を施して大気・気象データベースへと取り込むことになる。

また、個人情報に対する情報セキュリティの観点から、安全エリアと非安全エリアに分けたデータの管理を前提として設計を行うこととする。具体的には、臨床データを扱う「SS-MIX2匿名化データベース」（システムA）は個人情報を含むため、情報セキュリティ対策を施した安全エリア内でデータを取り扱うようにする。安全エリア内のデータは、SS-MIX2による匿名化処理を施すことなくエリア外には持ち出さないよう設計する。

環境データを扱う「各種外部環境データ管理システム」（システムB）では、個人情報が含まれないことから、個人情報を含まないことを前提として管理するエリアとして設計する。

同様に「診療-環境統合匿名化データウェアハウス」(システムC)も、匿名化された臨床データおよび環境データのみを取り扱うことから、個人情報を含まないことを前提として管理するエリアで動作するシステムとして設計する。

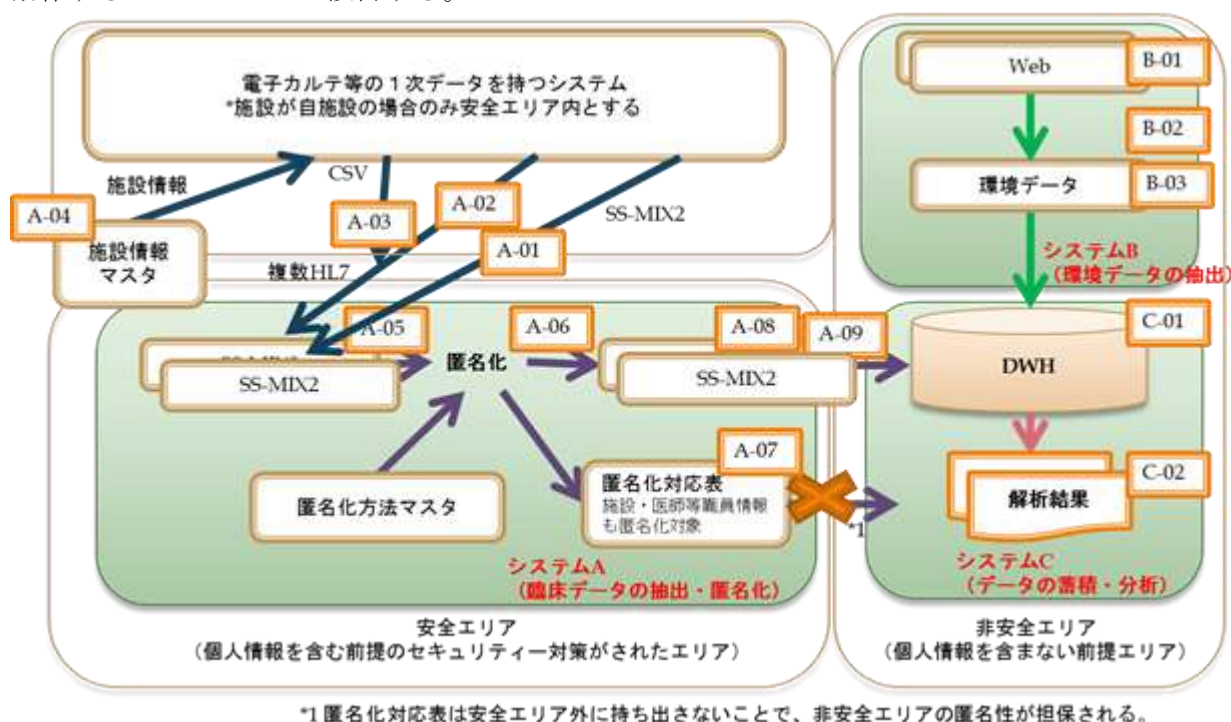


図3 個人情報保護のための安全エリアと非安全エリア別の設計

## (2) システムで取り込むデータの抽出元

病院情報システム内のデータとしては電子カルテMega-0akから抽出する医療データを、環境因子に関するデータとして環境省大気汚染物質広域監視システム（そらまめ君）、環境省花粉観測システム（はなこさん）、および気象庁気象観測データから取り込むデータを利用する。それぞれのデータ抽出元について以降に述べる。

### ① Mega-0ak

Mega-0akはNECの提供する電子カルテシステムである。カルテ入力作業の軽減、目的に応じた参照、EBM（根拠に基づく医療）のための情報蓄積などを支援するものであり、各種診療情報（オーダー、バイタル、検査結果各種レポート）、診察記事や経過表、検査画像、検査結果報告書、処方箋などのデータを蓄積している。



図 4 Mega-Oakの画面例 (NECホームページ<http://www.nec.co.jp/medsq/solution/hr/> より)

## ② 環境省大気汚染物質広域監視システム (そらまめ君)

そらまめ君は環境省の運営する、全国の大気汚染状況について、24時間の情報提供を行うサイトである。データとしては、1時間ごとの大気汚染測定結果、および 光化学オキシダント注意報・警報発令情報の最新 1 週間のデータを地図や表でみることができる。



図 5 そらまめ君の画面 (環境省 そらまめ君ホームページより<http://soramame.taiki.go.jp/>)

## ③ 環境省花粉観測システム (はなこさん)

はなこさんは環境省の運営する、花粉飛散状況を地図と表、グラフで見ることができるサイトである。花粉飛散データは観測地点から自動送信されており、毎時35分頃にホームページが更新される。花粉飛散状況は、花粉自動測定器により計測された、1時間平均の花粉数(個/m<sup>3</sup>)で表示している。なお留意点として、雪、黄砂、火山灰等が発生した際には数値が高く計測される場合がある。



図6 はなこさんの画面（環境省花粉観測システムホームページより <http://kafun.taiki.go.jp/>）

#### ④ 気象庁気象観測データ

気象庁の観測している現在・過去の気象観測データ、気象衛星画像、天気図などを、地図やグラフ、表の形式で見ることができるサイトである。各地の気温、降水量、風などのデータが長年にわたり全国で蓄積されている。



図7 気象庁気象観測データの画面

#### (3) 臨床データの匿名化方法

臨床データは、「厚生労働省電子的診療情報交換推進事業」（SS-MIX: Standardized Structured Medical record Information eXchange）による標準化ストレージの形式を用いる。

SS-MIXは、電子カルテ・オーダエントリを中心として、調剤システム、臨床検査システム、放射線情報システム、PACS等、様々な医療情報を標準化ストレージというストレージツールに医療情

報を標準化された形式で格納・蓄積するためのパッケージウェアの開発、配付などを行う取り組みである。

特徴としては、あらゆる医療施設で利用可能な設計であること、ハードウェア以外の初期投資が不要でソフトウェア保守等のコストが抑制されていること、特定の企業やベンダーの技術・製品に依存しない仕組みであること、病院情報システムに関する知識・スキルがあれば、特別な教育・研修を行わずとも誰もが理解しやすい単純な構造を採用していること、があげられる。

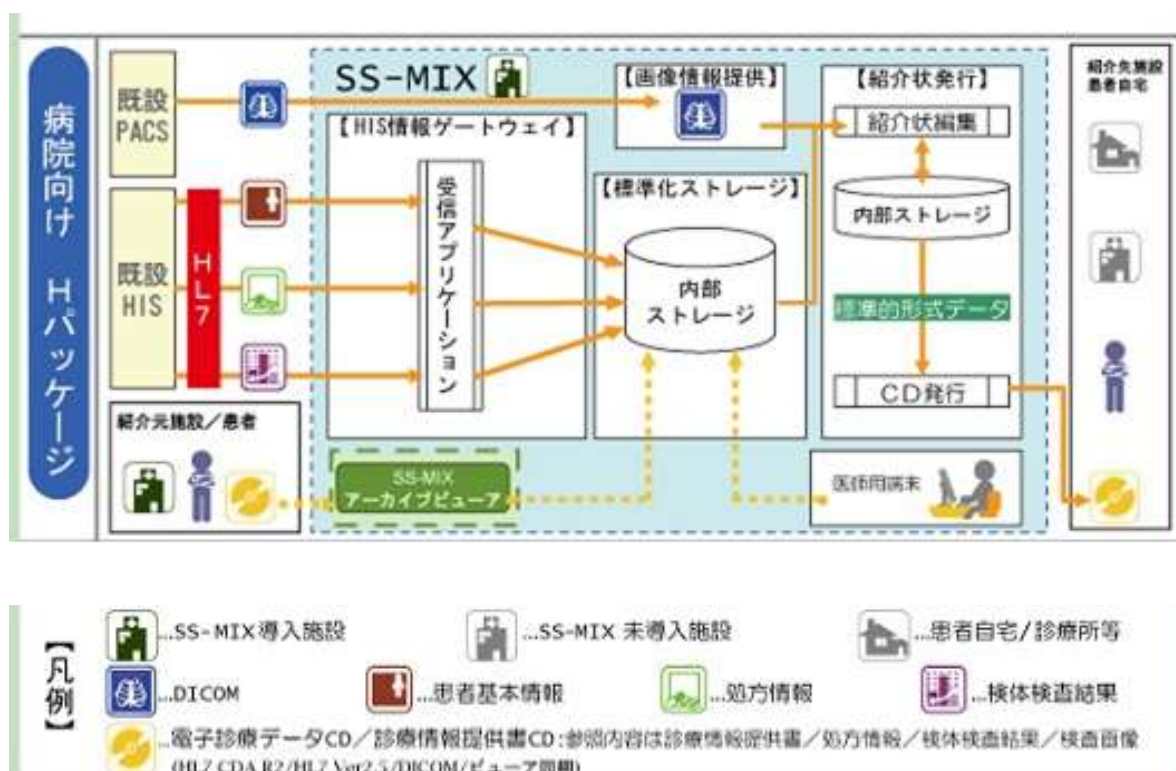


図 8 SS-MIXによる標準化ストレージの利用イメージ  
(SS-MIX普及推進コンソーシアムのHPより引用)

#### 4. 結果・考察

##### (1) 諸言

本研究の背景として、公害病についての研究が主な環境と疾病との問題点であった前世紀に比較し、現代においては温暖化をはじめとするグローバルな気候変動、PM2.5等の大気汚染物質に関する社会的な認知度が向上し、これら環境因子の健康影響に対する関心も上昇している。公害に対する理解から、各国は環境基準を策定している。適切な環境基準に基づき、環境行政・民間が共働して環境を維持することで、多くの先進国において重度の公害病は減少している。しかしながら、環境因子の変化に伴い人が罹患する疾病は変化していく。温暖化をはじめとし、地球規模での環境変化による健康影響を調査するための基盤づくりに、各国では環境因子のモニタリングがさかんに実施され、いままで環境因子の影響が指摘されてこなかった、あるいは経験的に指摘されていても、エビデンスが得られていなかった疾病に関して、環境因子と疾病構造に関する研究

が萌芽的に始まっている。我が国においても、全国を網羅し、一般公開されている環境ビッグデータが存在し、詳細な気象データやPM2.5を代表とする新たな環境因子が精密にモニタリングされている。そして徐々に新たな環境因子が疾病構造に影響を及ぼしていることが知られつつある。これら、従来研究の対象よりはるかに微細な環境因子の変動が、いかにわれわれの疾病構造に大きな影響をあたえているかを解析し、新たな知見を得るためには、いままで以上により大規模な症例群から、より詳細な臨床データを集めた、大きなデータベースが不可欠である。しかし、現在環境モニタリング結果と診療情報システムデータとを関連づけたデータ解析を行った事例は近年でもいまだに少数である。これは、最近まで診療情報が電子化されてこなかったこと、電子化されたデータへのフルアクセスが医師のみに制限されがちであること、診療データの記録目的が診療のためであり研究目的ではないことからデータに2次利用を行うためには様々な前処理、ノイズ処理が必要となることが障壁となってきたためである。また、研究の実現に情報科学の専門家と医師・医学者の協力が必要不可欠となるが両者の連携は専門性の違いから今までは行われることが少なかったなど、様々な技術的要因・社会的要因から環境モニタリング結果と診療情報の突合解析が行われてこなかったためである。

そこで本研究では、情報科学と医学の集学的な取り組みにおいて、環境データと診療情報とを関連付けたデータ解析を行い、今後複数の協力施設と連携してデータ解析を行うことを念頭に、環境因子との関連性が強い疾病候補を絞り込み、継続的な環境因子の変動と疾病構造の変化のモニタリングを可能にするインフラとして、病院情報システムのデータと、環境省が保有する環境データ（環境省大気汚染物質広域監視システムのデータ等）の統合及び解析を可能にする「共通システム」を構築した。以下に「共通システム」の全貌につき解説する。

## (2) 病院情報システムと環境ビッグデータ突合解析のための「共通システム」の概要

まず、病院情報システムと環境ビッグデータ突合解析のための「共通システム」の概要につき図 9 に示す。次いで図中①から⑥各項目について各々解説する。



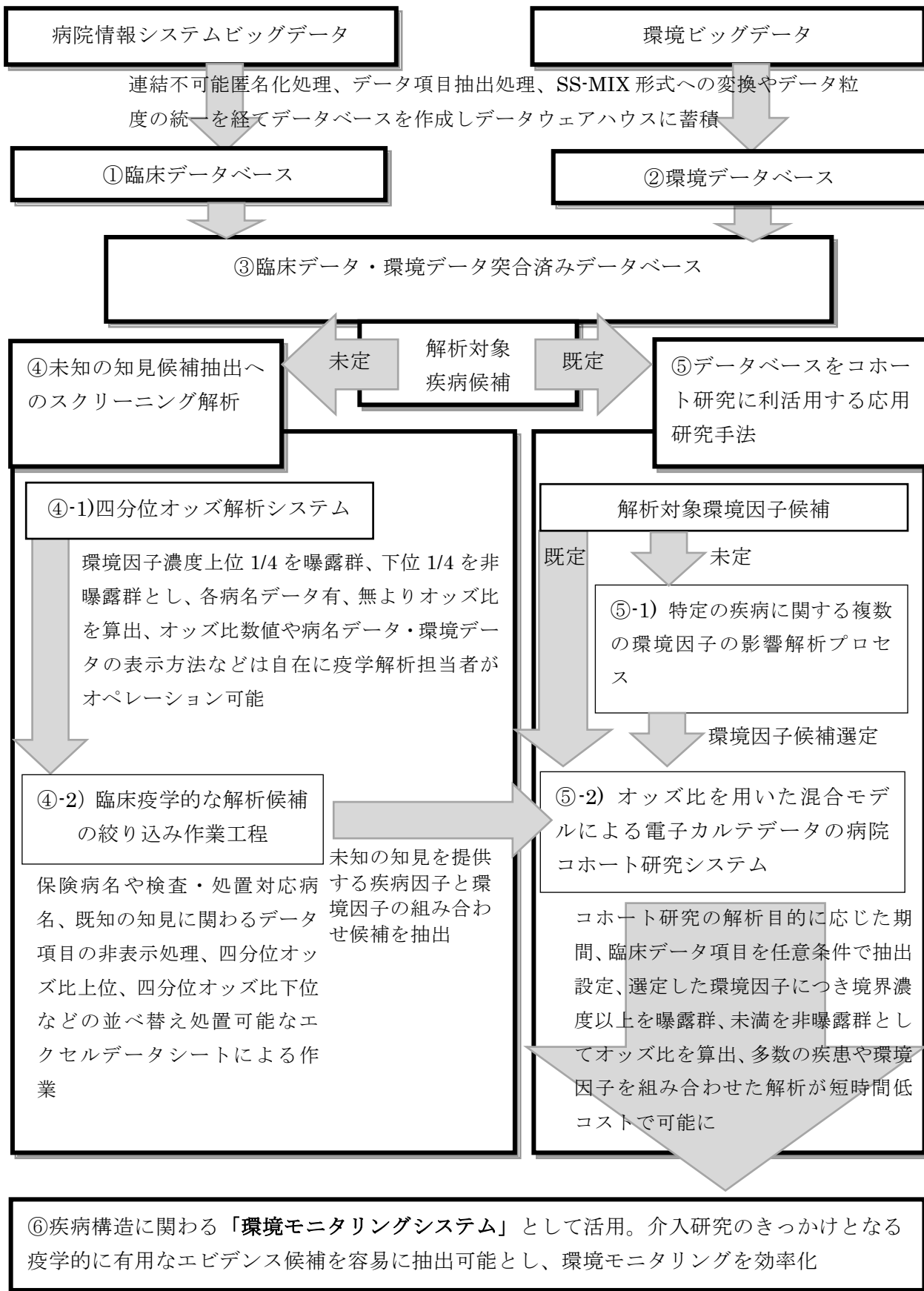


図9 病院情報システムと環境ビッグデータ突合解析のための「共通システム」の概要

## ① 病院情報システムビッグデータからの「臨床データベース」構築

臨床データについては、病院情報システム内のデータとしては電子カルテMega-Oakから抽出した。Mega-OakはNECの提供する電子カルテシステムである。カルテ入力作業の軽減、目的に応じた参照、EBM（根拠に基づく医療）のための情報蓄積などを支援するものであり、各種診療情報（オーダー、バイタル、検査結果各種レポート）、診察記事や経過表、検査画像、検査結果報告書、処方箋などのデータを蓄積している。医療データを病院システムから抽出した実データは、匿名化処理を施された後CDに書き込まれる。匿名化処理では、患者番号はランダムな番号に置き換えることで匿名化を実施している。氏名、カナ氏名に関しては、空白にすることで匿名化した。次に、CDからデータを読み込んでHL7の形式に変換し、SS\_MIXデータベースに取り込む。このように本データベースは、「厚生労働省電子的診療情報交換推進事業」（SS-MIX：Standardized Structured Medical record Information eXchange）による標準化ストレージの形式を用いる。SS-MIXは、電子カルテ・オーダーエントリを中心として、調剤システム、臨床検査システム、放射線情報システム、PACS等、様々な医療情報を標準化ストレージというストレージツールに医療情報を標準化された形式で格納・蓄積するためのパッケージウェアの開発、配付などを行う取り組みである。特徴としては、あらゆる医療施設で利用可能な設計であること、ハードウェア以外の初期投資が不要でソフトウェア保守等のコストが抑制されていること、特定の企業やベンダーの技術・製品に依存しない仕組みであること、病院情報システムに関する知識・スキルがあれば、特別な教育・研修を行わずとも誰もが理解しやすい単純な構造を採用していること、があげられる。また、個人情報に対する情報セキュリティの観点から、安全エリアと非安全エリアに分けたデータの管理を前提として設計を行った。解析用に加工済みデータとしてデータベースに構成されたおもな項目は下記で、実際の操作画面は図10である。

表 1臨床データベースから解析用に条件付けて抽出可能なおもな臨床データ項目

全ケースで抽出			
年齢	性別	郡市までの居所	保険種別
初受診日	初診日	再診日	受診診療科
病名	発症日	病名終了日	
転帰（治癒、治療継続、中断、死亡）			

一部入力のあったケースで抽出			
検査項目	検査結果	検査日時	投薬内容
投薬日時	処置内容	処置日時	
診療計測データ（収縮期血圧、拡張期血圧、脈拍、体重など）			

「臨床データベース」の蓄積成果

国立国際医療研究センター

2011年～2016年

匿名化患者 572384件

匿名化処方 1290253件

匿名化病名 239112件

匿名化予約 608147件

五百山クリニック

2014年

非アレルギー性鼻炎（NAR）の患者データ

著名改善 130件

改善 88件

不変・中断 45件

悪化 1件

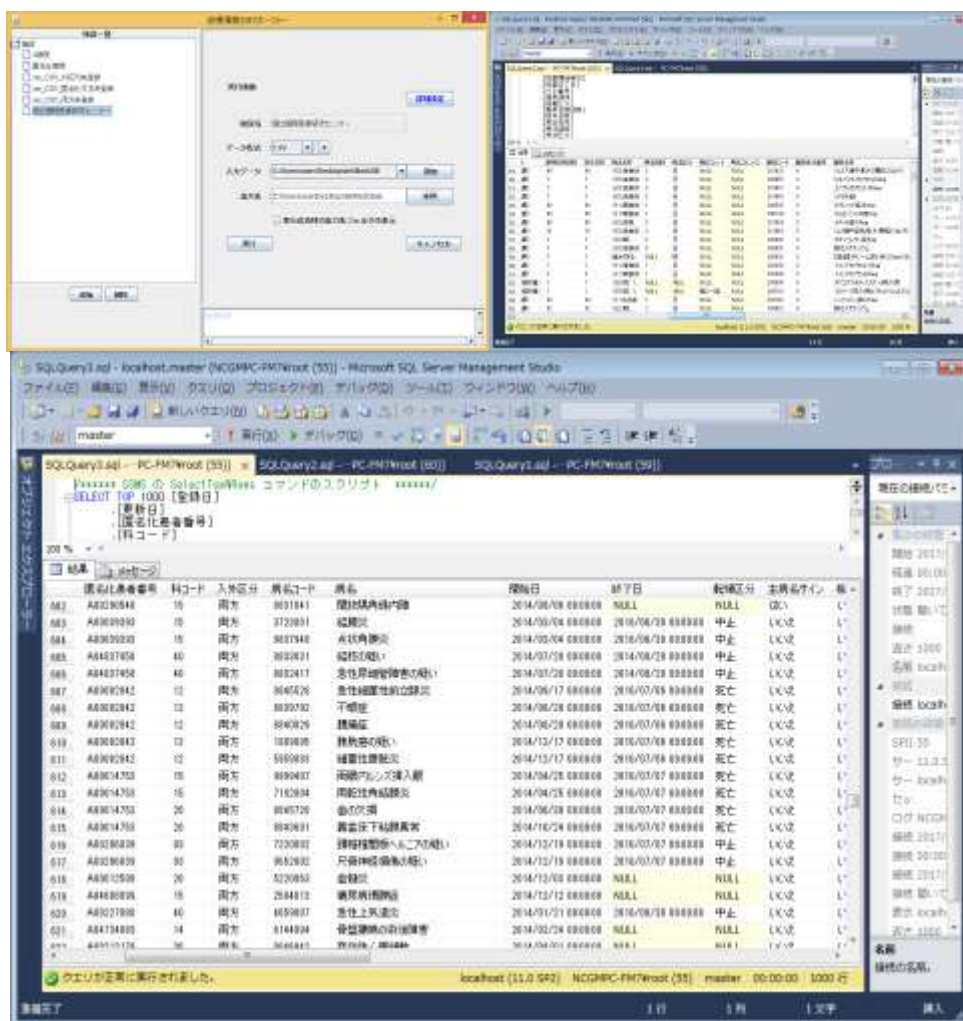


図 1 0 臨床データベースの操作画面

## ② 環境ビッグデータからの「環境データベース」構築

環境データについては、将来のリアルタイムの突合解析および継続的モニタリングを見据え、インターネット上の一般公開情報からデータをエクセル形式で取り込み、重複データの削除、時間単位データから日単位データへの集約などの処理を施して大気・気象データベースへと取り込んでいる。環境因子に関するデータとして環境省大気汚染物質広域監視システム（そらまめ君）、環境省花粉観測システム（はなこさん）、および気象庁気象観測データから取り込むデータを利用する。そらまめ君は環境省の運営する、全国の大気汚染状況について、24時間の情報提供を行うサイトである。データとしては、1時間ごとの大気汚染測定結果、および光化学オキシダント注意報・警報発令情報の最新1週間のデータを地図や表でみることができる。はなこさんは環境省の運営する、花粉飛散状況を地図と表、グラフで見ることができるサイトである。花粉飛散データは観測地点から自動送信されており、毎時35分頃にホームページが更新される。花粉飛散状況は、花粉自動測定器により計測された、1時間平均の花粉数（個/m<sup>3</sup>）で表示している。気象庁の観測している現在・過去の気象観測データ、気象衛星画像、天気図などを、地図やグラフ、表の形式で見ることができるサイトである。各地の気温、降水量、風などのデータが長年にわたり全国で蓄積されている。

環境データの項目の詳細については以下のデータをそれぞれの公開情報から抽出するよう設計した。これらの各データは、データウェアハウスの構築にあたっては、各データ抽出元によってサンプリング時間が異なるなどの問題がある。そのため、重複データの削除、時間単位データから日単位データへの集約などの処理などを施し、関連性分析を行いやすいようデータの質を揃えた上でデータウェアハウスに登録するようにした。環境データベースから解析用に条件付けて抽出可能なおもな環境データ項目を表2に示す。

表2 環境データベースから解析用に条件付けて抽出可能なおもな環境データ項目

気象データ			大気汚染物質や粒子状物質濃度	
平均現地圧	平均海面気圧	最低海面気圧	二酸化硫黄 (SO <sub>2</sub> )	一酸化窒素 (NO)
合計降水量	1時間あたりの最大降水量	10分間あたりの最大降水量	二酸化窒素 (NO <sub>2</sub> )	窒素酸化物 (NO <sub>x</sub> )
平均気温	最高気温	最低気温	一酸化炭素 (CO)	光化学オキシダント (OX)
気温差(最低気温と最高気温の温度差)			非メタン炭化水素 (NMHC)	メタン (CH <sub>4</sub> )
平均蒸気圧	平均湿度	最低湿度	全炭化水素濃度 (THC)	浮遊粒子状物質 (SPM)
平均風速	最大風速	最大瞬間風速	微細粒子状物質 (PM <sub>2.5</sub> )	花粉飛散数
日照時間	全天日射量			
降雪の深さの合計	最深積雪	平均雲量		

「環境データベース」の蓄積成果

- ・抽出対象地区 1,749件
- ・時間単位データ 15,372,000件 (92,064,000件2011年～2016年累積値)
- ・日単位データ 640,134件 (3,833,808件2011年～2016年累積値)

このように、環境データベースでは広くインターネット上からデータをエクセル形式で取り込み、重複データの削除、時間単位データから日単位データへの集約などの処理を施して環境データベースへと再構築する。取り込んだデータはSS-MIX側のネットワークに統合する。実際の操作画面を図 11 に示す。

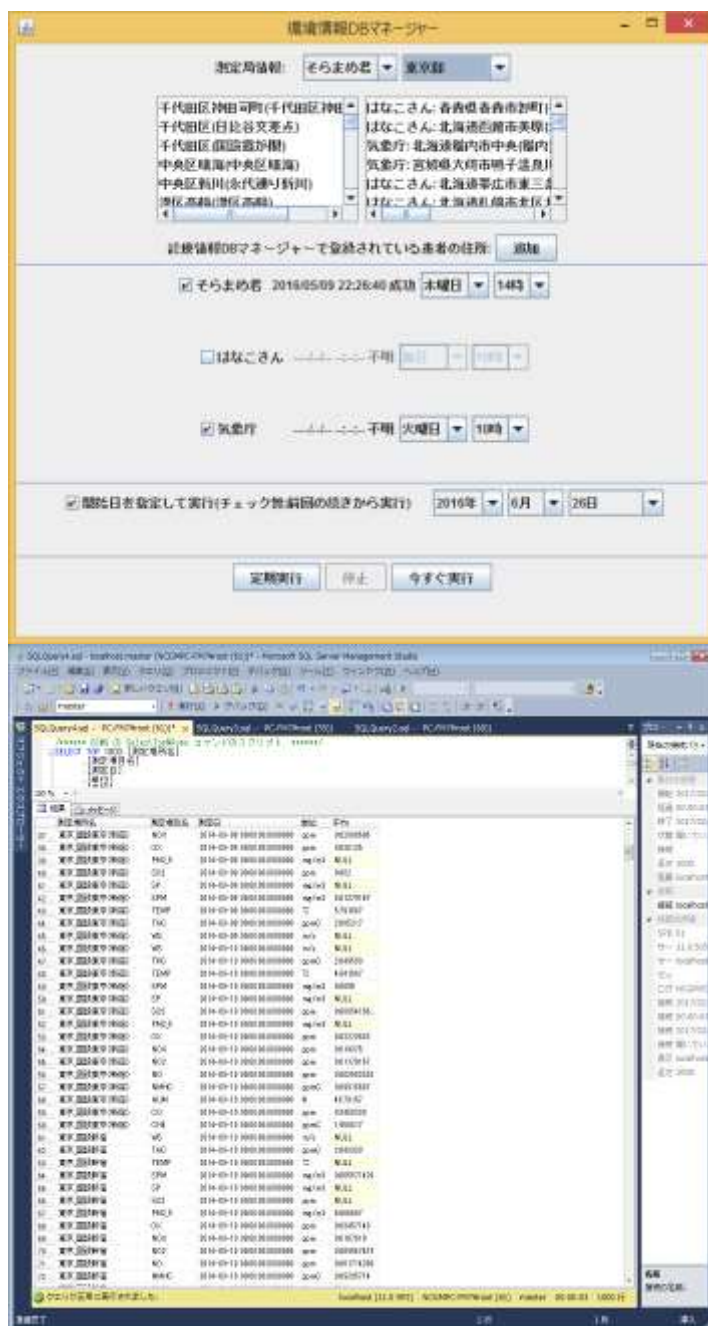


図 1 1 環境データベースの操作画面

### ③ 「臨床データ・環境データ突合済みデータベース」

「臨床データ・環境データ突合済みデータベース」では、データ取り込み・突合および蓄積のフローについては、2012年度に仕組み化した図2の成果を活用している。

臨床データおよび環境データを抽出し、統合したデータウェアハウスを構築するシステムの全体が完成した（平成27年度成果）。臨床データはSS-MIX2による匿名化処理を施すことで個人情報保護に対応し、環境データはデータ抽出元ごとに異なるデータの質を揃えることで関連性分析を行いやすいよう配慮している。システムによるデータの取り込み、蓄積が自動的に行われるよう、システムを構成する機能の自動化ツールをJavaによって作成した（平成27年度成果）。

これらのシステムを活用することによって、臨床データと環境データを効率的かつ継続的に収集し、以下に述べる臨床データと環境データの突合解析を効率的に行うことができるようになった。

「臨床データ・環境データ突合済みデータベース」の実例として、表3に60歳以上の糖尿病の罹患あり・なし、虚血発作発症あり・なしの疾病データと、環境因子を突合表示設定した際の、データテーブルの一部を例示する。本システムでは、このようなエクセルデータテーブルが、様々な条件で自動作成可能となっている。実際の操作画面を図12に示す。

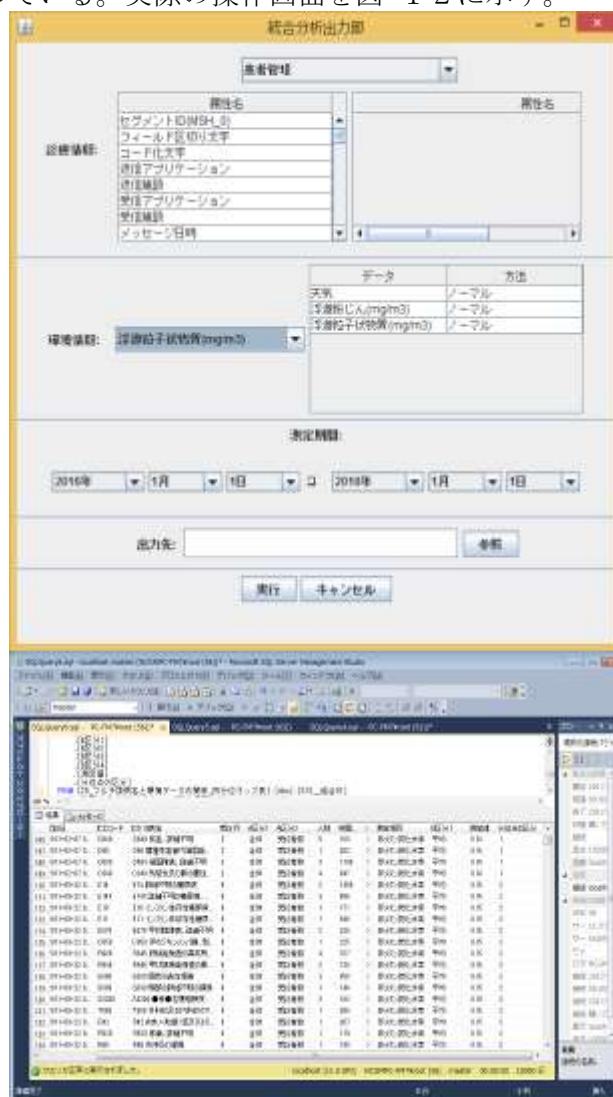


図12 臨床データ・環境データ突合済みデータベースの操作画面

表 3 「臨床データ・環境データ突合済みデータベース」の実例

日付	糖尿病発症あり	糖尿病発症なし	糖尿病発症あり	糖尿病発症なし	患者総計	休日名	休日区分	平均温度(℃)	PM2.5(μg/m <sup>3</sup> )	平均SO <sub>2</sub> (ppm)	平均NO(ppm)	平均NO <sub>2</sub> (ppm)
2011/01/02	1	4	5	23	33	年末年始	休日	5.7	6	0	0.002	0.014
2011/01/03	1	1	4	15	21	年末年始	休日	4.7	7	0	0.001	0.007

平均気温(℃)	最高気温(℃)	最低気温(℃)	降水量の合計(mm)	10分間降水量の最大(mm)	日照時間(時間)	全天日射量合計(J/m <sup>2</sup> )	最深積雪(cm)	降雪量合計(cm)	平均風速(m/s)	最大風速(m/s)	最大風速(m/s)	最大瞬間風速(m/s)	最大瞬間風速(m/s)	最多風向(16方位)	平均蒸気圧(hPa)	平均湿度(%)	最小相対湿度(%)	平均現地気圧(hPa)
7	11	4.1	0	0	8.4	11	0	0	2.2	4.6	北北西	8.4	北北西	北北西	4	41	27	1011.5
5.9	9.2	3.1	0	0	5.2	8.99	0	0	2.1	4.4	北北西	7.4	北北西	北北西	4.5	48	37	1011.5

## ④ 未知の知見候補抽出へのスクリーニング解析

平成27年までの成果により機能している「臨床データ・環境データ突合済みデータベース」を利用し、今年度の成果として、解析対象疾病候補が未定の場合を想定し、ある特定の環境因子に関連して疾病構造が変化する疾病候補を抽出するための、未知の知見候補抽出へのスクリーニング解析を行うシステムを開発・自動化した。

## ア 「四分位オッズ解析システム」(平成28年度自動化)

この解析システムでは、ある特定の環境因子を自由に設定でき、濃度上位1/4を曝露群、下位1/4を非曝露群とし、各病名データ有、無よりオッズ比を自動的に算出できる。四分位オッズ比の算出方法は表 4の通りである。オッズ比数値や病名データ・環境データの表示方法などは自在に疫学解析担当者がオペレーション可能となっている。

実際の操作画面を図 1 3 に示す。

表 4 四分位オッズ比

四分位オッズ比の算出方法  $\frac{a}{c} \div \frac{b}{d} = \frac{ad}{bc}$

	症例： 病名あり	対照： 病名なし
曝露群：環境因子 濃度 上位1/4	症例数 <i>a</i>	症例数 <i>c</i>
非曝露群：環境因子 濃度 下位1/4	症例数 <i>b</i>	症例数 <i>d</i>

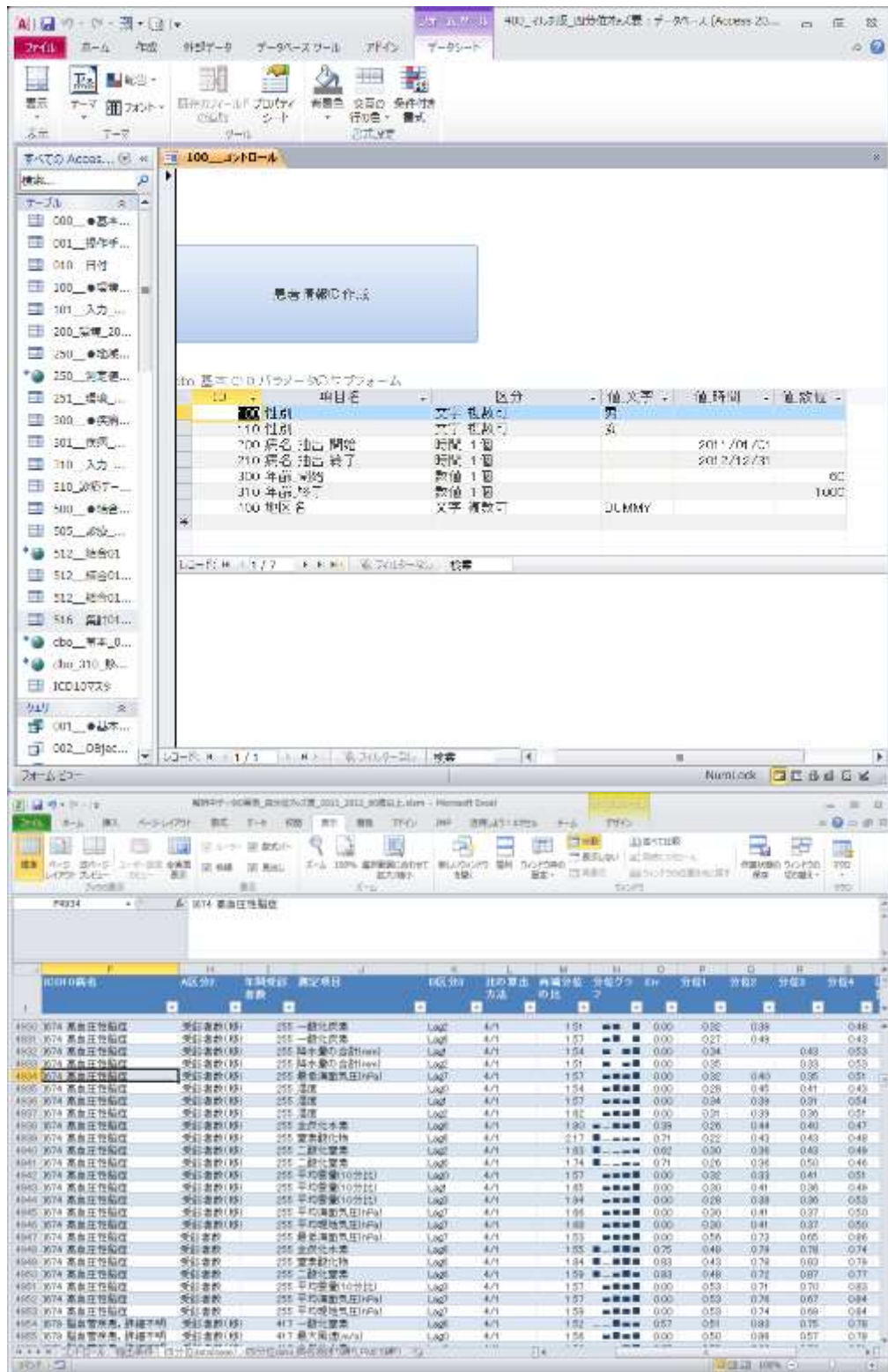


図 1 3 四分位オッズ解析システムの操作画面

イ 臨床疫学的な解析候補の絞り込み作業工程（平成28年度ソフトウェアとして統合）

実際に解析を担当する、生気象学者や、臨床疫学者が、知識と臨床経験により、保険病名や検査・処置対応病名、既知の知見に関わるデータ項目の非表示処理、四分位オッズ比上位、四分位オッズ比下位などの並べ替え処置を、自由に操作可能なエクセルデータシートにより行い、未知の知



見を提供する疾病因子と環境因子の組み合わせ候補を抽出する作業。この部分は自動化せず、マニュアルで操作可能なソフトウェアとなっている。この手法を確立するため、のプレスタディを兼ねた応用研究として、われわれ研究班が施行した疫学研究より、PM2.5に関する疾病候補解析工程画面の実例を図1-4に示す。

年	ICD コード	ICD10病名	年間受 診者数	測定項目	B区分3	比の算 出方法	両端分 位の比	分位グラ フ
2012	I802	I802 下肢のその他の深在血管の静脈炎	104	PM2.5	Lag0	4/1	2.39	■■■■■
2011	I469	I469 心停止, 詳細不明	131	PM2.5	Lag1	1/4	1.93	■■■■■
2012	I709	I709 全身性及び詳細不明のアテローム<	134	PM2.5	Lag0	4/1	1.89	■■■■■
2011	I638	I638 その他の脳梗塞	115	PM2.5	Lag2	1/4	1.80	■■■■■
2012	I693	I693 脳梗塞の続発・後遺症	163	PM2.5	Lag2	1/4	1.77	■■■■■
2012	I652	I652 頸動脈の閉塞及び狭窄	173	PM2.5	Lag0	4/1	1.74	■■■■■
2012	I252	I252 陳旧性心筋梗塞	146	PM2.5	Lag2	1/4	1.72	■■■■■
2012	I693	I693 脳梗塞の続発・後遺症	163	PM2.5	Lag2	1/4	1.72	■■■■■
2012	I638	I638 その他の脳梗塞	125	PM2.5	Lag0	4/1	1.69	■■■■■
2011	I259	I259 慢性虚血性心疾患, 詳細不明	130	PM2.5	Lag0	4/1	1.68	■■■■■
2012	I269	I269 急性肺性心の記載のない肺塞栓症	108	PM2.5	Lag1	1/4	1.64	■■■■■
2011	I269	I269 急性肺性心の記載のない肺塞栓症	111	PM2.5	Lag1	4/1	1.63	■■■■■
2011	I259	I259 慢性虚血性心疾患, 詳細不明	130	PM2.5	Lag0	4/1	1.62	■■■■■
2012	I633	I633 脳動脈の血栓症による脳梗塞	118	PM2.5	Lag0	4/1	1.54	■■■■■
2012	I652	I652 頸動脈の閉塞及び狭窄	173	PM2.5	Lag2	1/4	1.53	■■■■■
2012	I269	I269 急性肺性心の記載のない肺塞栓症	108	PM2.5	Lag2	4/1	1.53	■■■■■
2012	I693	I693 脳梗塞の続発・後遺症	163	PM2.5	Lag0	4/1	1.52	■■■■■

図1-4 四分位オッズ解析システムを用いた臨床疫学的な解析候補の絞り込み作業工程の具体例

#### ⑤ データベースをコホート研究に利活用する応用研究手法

臨床データ・環境データ突合済みデータベースを用い、解析対象疾病候補が既に定まっている場合の、応用研究手法に関わるプロセスと、本研究で今年度作成、自動化したシステムにつき、解説する。

##### ア 特定の疾病に関する複数の環境因子の影響解析プロセス

既に解析対象環境因子が決まっている場合は直接2)に進む。解析対象環境因子候補が未定の場合、まず、臨床データ・環境データ突合済みデータベースに付帯させた解析機能を用い、通常の疫学研究手法に沿い偏相関分析・重回帰分析などの手法で解析対象疾患との影響を調べたい環境因子を選定・抽出する。

##### イ 「オッズ比を用いた混合モデルによる電子カルテデータの病院コホート研究システム（平成28年度自動化）

本システムでは、コホート研究の解析目的に応じた任意の期間、「臨床データ・環境データ突合済みデータベース」より臨床データ項目や環境因子を任意条件で抽出設定、選定した環境因子につき境界濃度以上を曝露群、未満を非曝露群としてオッズ比をくりかえし算出、多数の疾患や環境因子を組み合わせた解析が短時間低コストで可能になる。昨年度までに上記一連の解析システムを作成していたが、本年度は今後の無償配布を念頭にこのシステムを統合し、表5、図1-5、図1-6にPM2.5濃度境界と虚血発作のおこりやすさを解析した事例(1)に関して示すような作業工程の全自動化作業を施行した（平成28年度完成）。実際の解析ソフト画面を図1-7に解析結果出力画面の実例を図1-8に示す。この解析結果では、温暖化の疾病構造への影響を念頭に温暖時期の各気温境界でのくも膜下出血のおこりやすさへの影響をスクリーニング解析している。

表 5 カウントのイメージとオッズ比の計算式

PM2.5	虚血発作	
	有	無
設定した境界値以上	A	B
設定した境界値未満	C	D

オッズ比 =  $\frac{(A/B)}{(C/D)}$

		糖尿病 (有)				(He/Hc) / (Le/Lc)	
		気温	脳心梗塞(有)	脳心梗塞(無)			
10	μg/m <sup>3</sup>	Lag2 25-	977	83	2308	175	0.89
11	μg/m <sup>3</sup>	Lag2 25-	852	208	2010	473	0.96
12	μg/m <sup>3</sup>	Lag2 25-	766	294	1769	714	1.05
13	μg/m <sup>3</sup>	Lag2 25-	684	376	1552	931	1.09
14	μg/m <sup>3</sup>	Lag2 25-	598	462	1357	1126	1.07
15	μg/m <sup>3</sup>	Lag2 25-	488	572	1078	1405	1.11
16	μg/m <sup>3</sup>	Lag2 25-	424	636	897	1586	1.18
17	μg/m <sup>3</sup>	Lag2 25-	361	699	765	1718	1.16
18	μg/m <sup>3</sup>	Lag2 25-	300	760	616	1867	1.20
19	μg/m <sup>3</sup>	Lag2 25-	245	815	507	1976	1.17
20	μg/m <sup>3</sup>	Lag2 25-	222	838	436	2047	1.24
21	μg/m <sup>3</sup>	Lag2 25-	189	871	391	2092	1.16
22	μg/m <sup>3</sup>	Lag2 25-	173	887	341	2142	1.23
23	μg/m <sup>3</sup>	Lag2 25-	137	923	275	2208	1.19
24	μg/m <sup>3</sup>	Lag2 25-	128	932	253	2230	1.21
25	μg/m <sup>3</sup>	Lag2 25-	86	974	183	2300	1.11
26	μg/m <sup>3</sup>	Lag2 25-	75	985	163	2320	1.08
27	μg/m <sup>3</sup>	Lag2 25-	44	1016	101	2382	1.02
28	μg/m <sup>3</sup>	Lag2 25-	33	1027	76	2407	1.02
29	μg/m <sup>3</sup>	Lag2 25-	31	1029	74	2409	0.98
30	μg/m <sup>3</sup>	Lag2 25-	29	1031	72	2411	0.94

図 1 5 オッズ比を用いた混合モデルによる電子カルテデータの病院コホート研究システムによる解析中データの実例

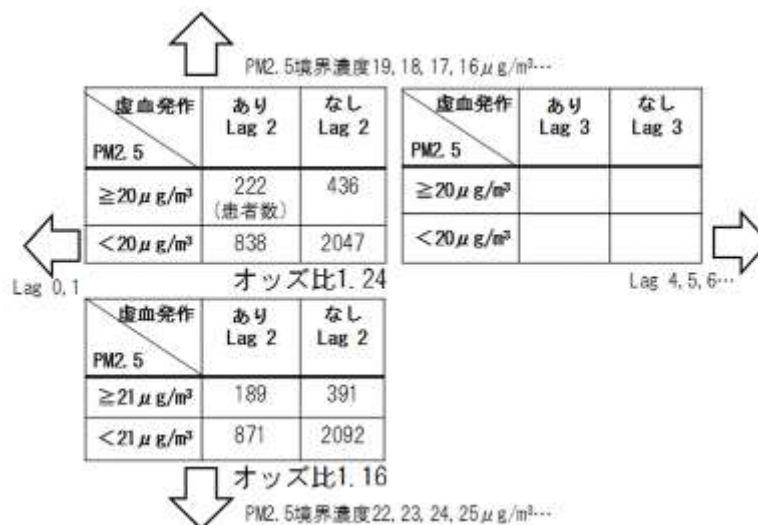


図 1 6 PM2.5濃度と曝露・発症タイムラグの解析方法

The screenshot shows the Microsoft Access 2010 interface. The main window displays a table with the following columns: 性別グループ (Gender Group), 既往症グループ (Pre-existing Condition Group), 発症病名グループ (Onset Disease Name Group), 環境項目名1 (Environment Item Name 1), 環境項目名2 (Environment Item Name 2), 年齢区分 (Age Group), 処置状態 (Treatment Status), and コメント (Comment). The table contains 47 rows of data, each representing a different medical condition or symptom.

性別グループ	既往症グループ	発症病名グループ	環境項目名1	環境項目名2	年齢区分	処置状態	コメント
1	両方	既往症	1.くも膜下出血	MAX_TEMP	MAX_TEMP	65-	20170215_30
2	両方	既往症	1.くも膜下出血	MAX_TEMP	MAX_TEMP	全年齢	20170215_30
3	両方	既往症	J13 肺炎レンサ球菌による肺炎	温度	PM2.5	60-	69.[環境省向]
4	両方	既往症	J810 慢性肺炎(アレルギー性)	温度	PM2.5	60-	68.[環境省向]
5	両方	既往症	T793 外傷後の創傷感染症, 0	温度	PM2.5	60-	67.[環境省向]
6	両方	既往症	I839 潰瘍又は炎を伴わない	温度	PM2.5	60-	66.[環境省向]
7	両方	既往症	D876 口腔及び消化器の性状	温度	PM2.5	60-	65.[環境省向]
8	両方	既往症	R100 急性腹症	温度	PM2.5	60-	64.[環境省向]
9	両方	既往症	D699 出血性病態, 詳細不明	温度	PM2.5	60-	63.[環境省向]
10	両方	既往症	E702 尿素サイクル代謝障害	温度	PM2.5	60-	62.[環境省向]
11	両方	既往症	I499 不整脈, 詳細不明	温度	PM2.5	60-	60.[環境省向]
12	両方	既往症	S065 外傷性硬膜下出血	温度	PM2.5	60-	59.[環境省向]
13	両方	既往症	N300 急性膀胱炎	温度	PM2.5	60-	58.[環境省向]
14	両方	既往症	G20 パーキンソン病	温度	PM2.5	60-	57.[環境省向]
15	両方	既往症	A049 細菌性腸管感染症, 詳細不明	温度	PM2.5	60-	55.[環境省向]
16	両方	既往症	I252 陳旧性心筋梗塞	温度	PM2.5	60-	54.[環境省向]
17	両方	既往症	R521 慢性難治性疼痛	温度	PM2.5	60-	53.[環境省向]
18	両方	既往症	I252 陳旧性心筋梗塞	温度	オキシダント	60-	52.[環境省向]
19	両方	既往症	K85 急性肝炎	温度	PM2.5	60-	52.[環境省向]
20	両方	既往症	D27 卵巣の良性新生物	温度	PM2.5	60-	51.[環境省向]
21	両方	既往症	I749 詳細不明の動脈の血栓症	温度	オキシダント	60-	51.[環境省向]
22	両方	既往症	H811 良性発作性のまぶし眩暈	温度	PM2.5	60-	50.[環境省向]
23	両方	既往症	E99 その他及び詳細不明の死	温度	オキシダント	60-	49.[環境省向]
24	両方	既往症	S099 頭部の詳細不明の損傷	温度	PM2.5	60-	49.[環境省向]
25	両方	既往症	E875 高カウム<K>血症	温度	PM2.5	60-	48.[環境省向]
26	両方	既往症	H46 視神経炎	温度	オキシダント	60-	48.[環境省向]
27	両方	既往症	M701 筋肉痛	温度	オキシダント	60-	47.[環境省向]

図 1 7 オッズ比を用いた混合モデルによる電子カルテデータの病院コホート研究システムの解析画面

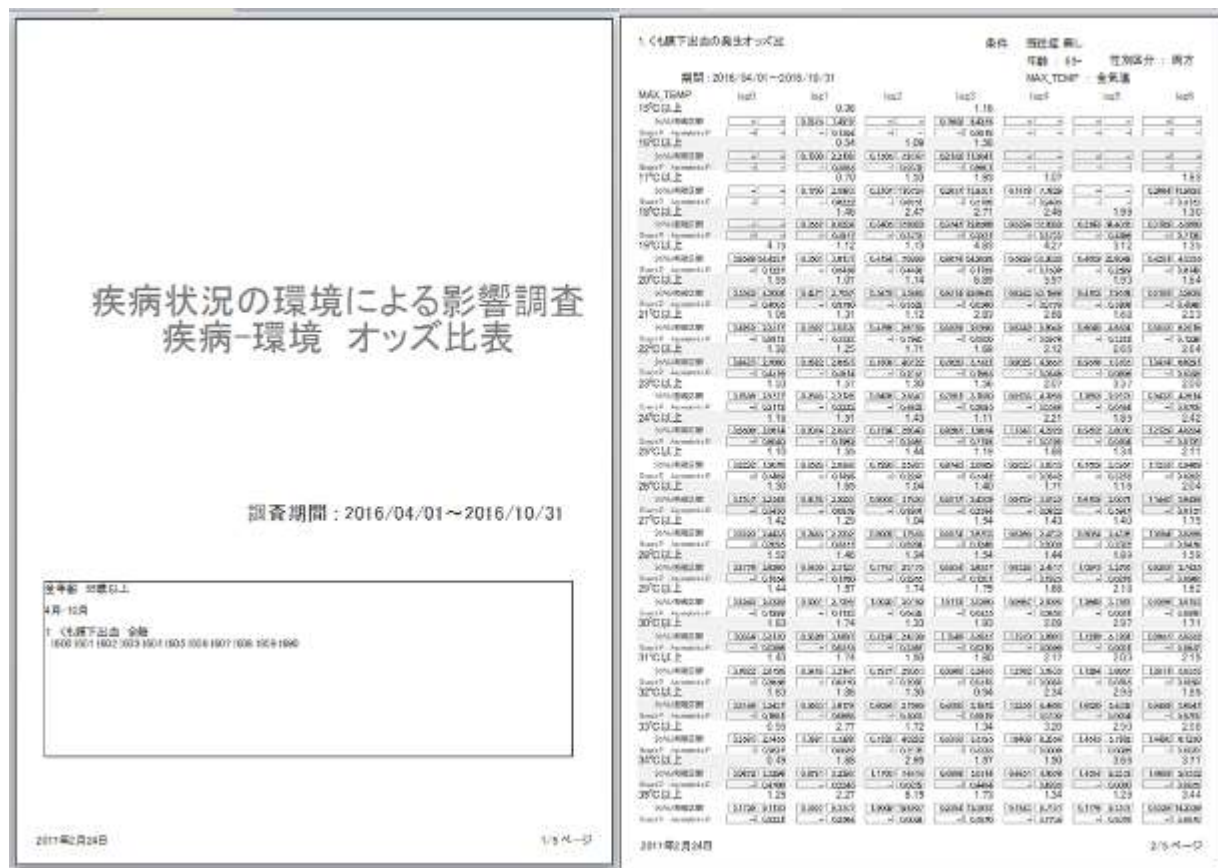


図 1 8 オッズ比を用いた混合モデルによる電子カルテデータの病院コホート研究システムの解析結果出力画面の実例

⑥ 「共通システム」(①から⑤で構成される一連の解析システムの総称)から「環境モニタリングシステム」へ

共通システムを中長期に継続的に運用し、今後、複数施設に導入・連携していくことで、多岐にわたって応用可能な、疾病構造解析に関わる「環境モニタリングシステム」として活用することができる。また、介入研究のきっかけとなる疫学的に有用なエビデンス候補が容易に抽出可能となり、環境モニタリングとその応用が飛躍的に効率化され、イノベーションまでのスピードアップが期待される。実際、本研究班でも先述した「オッズ比を用いた混合モデルによる電子カルテデータの病院コホート研究システム(平成28年度自動化)」開発にむけて、実装しながら改良を繰り返すスパイラルモデルによる開発の過程で、解析候補抽出のためのプレスタディを積み重ねながら、II部で述べる多数の解析結果を得た。このような解析結果を生気象学者や臨床疫学者、統計学の専門家が集学的に検討することで、介入研究のきっかけとなる疫学的に有用なエビデンス候補抽出が効率的に抽出、必要に応じて中長期のモニタリング候補に設定することも可能であるため、将来的なモバイル端末を介したリアルタイムの健康天気予報など、イノベーションに貢献する基盤となると考える。

## 5. 本研究により得られた成果

### (1) 本研究班が行った具体的な応用解析例

#### ① 諸言

まず医学研究のエビデンスレベルについて日本医療機能評価機構が運営する医療情報サービス事業 Minds の概念図より図に示す。縦軸に解析の論理的正確性を、横軸に解析対象症例数を示している。解析対象症例数が増え、かつ、解析の論理的正確性が増してこの図で右上にあがるほど、高いエビデンスが得られるものの、何年もの時間と多くのマンパワーを要し、高額な予算が必要となり、解析対象が限られてしまう。このため、ほんとうに疫学データを必要とする、少数例だが重篤で難治な疾患や、軽症だが慢性で生活に支障をきたす疾患を、さまざまな条件で解析するには、短時間ローコストで解析できる方法論が希求されている。その候補として、この図のように、情報学的手法を駆使して、いままでの紙カルテベースでは収集不可能であった、膨大な症例データを電子カルテビッグデータから抽出、解析する方法が、今年度完成した「共通システム」である。このシステムでは、解析の論理的正確性はIVのままだが、飛躍的に解析対象症例数を増やすことで、まずはこの概念図の右に移動して、環境情報学的な側面から、いくつものエビデンス候補となる新知見を見出せば、勝算ある分野として、III、IIのレベルの解析が行われるための基礎資料となり、希少難治な疾患の疫学研究も、進展しやすくなることが期待される。本研究では、環境因子データと、疾病情報を含んだ電子カルテビッグデータを突合解析し、解析過程を含め学術誌に発表することで、この方法論で得た疫学データを社会に還元している<sup>(1),(2)</sup>。

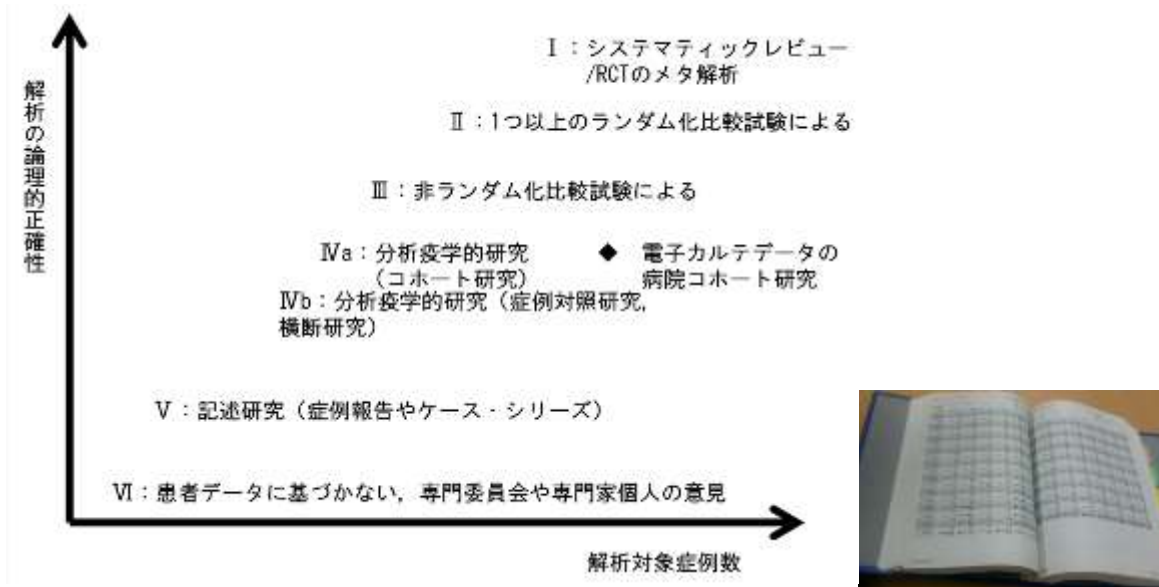


図19 医学研究のエビデンスレベル(日本医療機能評価機構が運営する医療情報サービス事業 Minds より引用一部改編)

#### ② 「共通システム」を用いて現在進行中の応用解析成果の実例

スパイラルメソッドによる「共通システム」開発を通じ、実際の解析を施行しながら「共通システム」を自動化する作業工程を経て、このような解析結果を電話帳ほどの質量となる合計890解析候補疾患に関して12か月前後のわずかな解析期間で得ることができている。今後、さらに糖尿病などの内分泌疾患、前項で実例に示した、くも膜下出血などの救急疾患、耳鼻咽喉科領域、循環器疾患、消化器疾患、呼吸器疾患領域などの専門医にこの結果を提供・連携し、解析内容を

臨床疫学的に深化させる計画である。解析候補疾患の一部を表 1に、「共通システム」を用いて現在進行中の応用解析成果の一部を実例として示す。現在進行中の応用解析成果では、表 2に示す、喘息既往ありの患者群で25℃以上の夏日にPM2.5曝露と嚥下性気管支炎・肺炎の発症オッズが濃度依存性に上昇する現象や、10℃未満の寒冷環境下でのPM2.5曝露と高齢糖尿病患者の虚血発作発症にPM2.5濃度依存性が認められている。また、現在本研究班では、温暖化の疾病構造への影響を念頭に温暖時期のくも膜下出血のおこりやすさへの影響をスクリーニング解析している。粗データではあるが、温暖時期の最高気温と曝露後の全年齢糖尿病既往のない患者群におけるくも膜下出血発症との関連解析において、高気温に曝露された数日後、とくに2日後に最高気温上昇に依存して、有意にくも膜下出血発症が増加する新知見を得ており(図)、今後詳細な解析へ進みたい。誤嚥性肺炎や、虚血発作(脳梗塞および心筋梗塞)、くも膜下出血はいずれも生命予後が悪く、治療に高額な医療費を要し、回復後も後遺障害により長期の介護を要する疾患であることから、今後の詳細な解析が有用と考えている。

表 1 「共通システム」を用いて抽出された解析対象候補疾患

発症から受診までのタイムラグ	環境因子曝露から発症までのタイムラグ	解析対象候補疾患	解析候補環境因子
少ない	直後から数日	虚血発作 くも膜下出血 重篤な不整脈 重症呼吸器感染症 重症喘息発作 緑内障発作 重症片頭痛発作 虫垂炎	温暖条件、寒冷条件、PM2.5 PM2.5 大気汚染物質、気温、湿度 気圧変動、温度差 気圧 気圧、降水量、気温差 日照時間
1週以内程度	少ない	非アレルギー性鼻炎 アレルギー性鼻炎 アレルギー性結膜炎 蕁麻疹 神経症 季節性のある精神病 高血圧症 関節炎	気温差、降水量、気温差、日射量 花粉、大気汚染物質 花粉、大気汚染物質 気圧、温暖条件、寒冷条件 風速、気温差 日照時間 温度、湿度、風速 湿度、温度、気圧

表 2 喘息患者群の夏日における PM2.5 曝露と嚥下性気管支炎・肺炎の発症との関連

喘息既往ありの患者群で 25℃以上の夏日にPM2.5曝露と嚥下性気管支炎・肺炎の発症				
PM2.5濃度 ( $\geq \mu\text{g}/\text{m}^3$ )	lag0	lag1	lag2	lag3
10	0.95	1.15	1.93	0.52
11	0.87	1.38	0.96	1.04
12	0.97	1.22	1.04	0.93
13	0.99	1.03	0.97	0.95
14	0.87	0.95	0.99	0.93
15	0.87	1.00	0.92	0.76
16	1.03	1.08	0.87	0.85
17	1.17	1.15	0.91	0.82
18	1.03	1.35	0.88	0.70
19	1.12	1.56	1.02	0.85
20	1.15	1.68	1.11	0.82
21	1.17	1.71	1.25	0.86
22	1.32	1.53	1.35	1.02
23	1.32	1.53	1.07	1.07
24	1.34	1.57	1.18	0.92
25	1.41	1.29	1.53	1.19
26	1.54	1.09	1.23	1.23
27	1.52	1.43	1.73	1.18
28	1.74	0.85	2.27	0.75
29	2.35	1.15	2.46	0.78
30	2.84	1.70	2.60	0.80

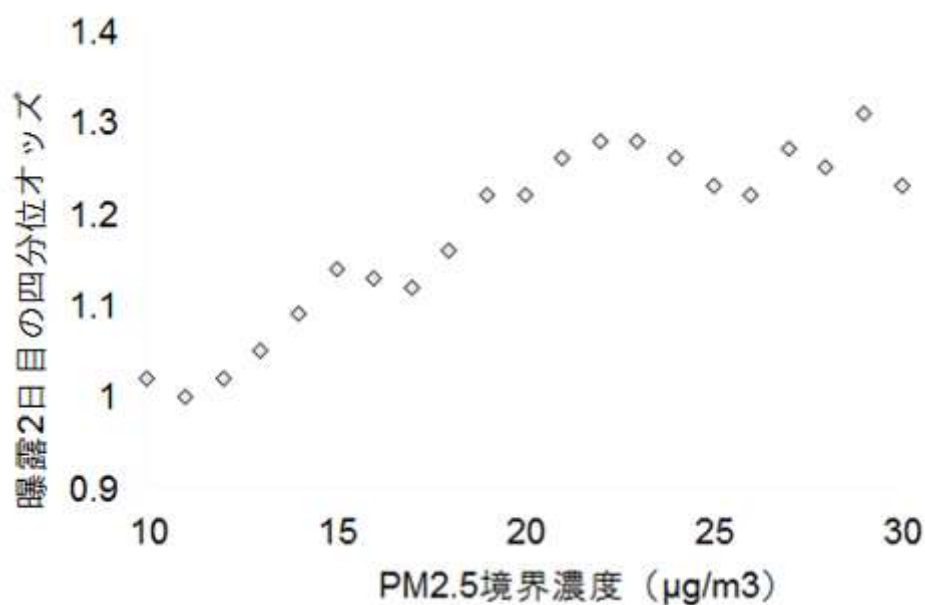


図 20 寒冷環境下での PM2.5 曝露と糖尿病高齢者における曝露 2 日目の虚血発作発症四分位オッズ比の PM2.5 濃度依存性

図 2 1 温暖時期の最高気温と糖尿病既往のない全年齢患者群におけるくも膜下出血発症との関連

期間：2016/04/01～2016/10/31		MAX_TEMP：全気温						
MAX_TEMP	lag0	lag1	lag2	lag3	lag4	lag5	lag6	
<b>25℃以上</b>	1.35	1.67	1.20	1.43	1.43	1.61	1.94	
95%信頼区間	0.9765 1.8591	1.1853 2.3561	0.8732 1.6625	1.0284 1.9810	1.0358 1.9829	1.1503 2.2659	1.3666 2.7435	
Exact P Asymptotic P	- 0.0695	- 0.0034	- 0.2565	- 0.0334	- 0.0298	- 0.0056	- 0.0002	
<b>26℃以上</b>	1.41	1.60	1.14	1.51	1.32	1.37	1.94	
95%信頼区間	1.0296 1.9186	1.1601 2.2028	0.8393 1.5569	1.1011 2.0682	0.9681 1.8017	0.9987 1.8663	1.4049 2.6899	
Exact P Asymptotic P	- 0.0320	- 0.0041	- 0.3962	- 0.0105	- 0.0792	- 0.0509	- 0.0001	
<b>27℃以上</b>	1.60	1.55	1.21	1.58	1.29	1.42	1.82	
95%信頼区間	1.1759 2.1813	1.1351 2.1151	0.8890 1.6422	1.1586 2.1589	0.9521 1.7610	1.0411 1.9312	1.3326 2.4866	
Exact P Asymptotic P	- 0.0028	- 0.0058	- 0.2269	- 0.0039	- 0.0995	- 0.0267	- 0.0002	
<b>28℃以上</b>	1.58	1.52	1.36	1.35	1.36	1.51	1.65	
95%信頼区間	1.1633 2.1518	1.1148 2.0601	1.0025 1.8581	0.9910 1.8354	0.9998 1.8513	1.1075 2.0459	1.2140 2.2434	
Exact P Asymptotic P	- 0.0035	- 0.0080	- 0.0482	- 0.0571	- 0.0503	- 0.0090	- 0.0014	
<b>29℃以上</b>	1.68	1.62	1.47	1.24	1.38	1.51	1.68	
95%信頼区間	1.2336 2.2934	1.1864 2.2111	1.0712 2.0124	0.9045 1.7097	1.0046 1.8909	1.1018 2.0627	1.2285 2.2896	
Exact P Asymptotic P	- 0.0010	- 0.0024	- 0.0170	- 0.1796	- 0.0468	- 0.0103	- 0.0011	
<b>30℃以上</b>	1.82	1.38	1.54	1.45	1.47	1.82	1.75	
95%信頼区間	1.3272 2.5092	0.9867 1.9286	1.1096 2.1381	1.0478 2.0189	1.0576 2.0517	1.3201 2.4959	1.2685 2.4097	
Exact P Asymptotic P	- 0.0002	- 0.0598	- 0.0098	- 0.0251	- 0.0220	- 0.0002	- 0.0006	
<b>31℃以上</b>	1.61	1.39	1.55	1.43	1.50	1.72	1.92	
95%信頼区間	1.1402 2.2677	0.9690 2.0048	1.0945 2.1986	1.0090 2.0267	1.0555 2.1316	1.2224 2.4201	1.3790 2.6656	
Exact P Asymptotic P	- 0.0068	- 0.0734	- 0.0136	- 0.0444	- 0.0237	- 0.0019	- 0.0001	
<b>32℃以上</b>	1.78	1.53	1.57	1.26	1.38	1.87	1.65	
95%信頼区間	1.2183 2.5973	1.0321 2.2825	1.0574 2.3384	0.8212 1.9278	0.9058 2.0979	1.2775 2.7237	1.1223 2.4345	
Exact P Asymptotic P	- 0.0029	- 0.0343	- 0.0254	- 0.2914	- 0.1341	- 0.0012	- 0.0110	
<b>33℃以上</b>	1.50	1.66	1.80	1.58	1.83	2.15	2.35	
95%信頼区間	0.9287 2.4225	1.0473 2.6239	1.1230 2.8693	0.9762 2.5467	1.1757 2.8458	1.4011 3.2923	1.5203 3.6253	
Exact P Asymptotic P	- 0.0974	- 0.0310	- 0.0143	- 0.0627	- 0.0074	- 0.0005	- 0.0001	
<b>34℃以上</b>	1.05	1.02	1.97	1.90	1.06	1.91	3.29	
95%信頼区間	0.4903 2.2351	0.4786 2.1819	1.0678 3.6447	1.0529 3.4235	0.4977 2.2689	0.9737 3.7496	1.9283 5.6063	
Exact P Asymptotic P	- 0.9059	- 0.9553	- 0.0300	- 0.0331	- 0.8753	- 0.0598	- 0.0000	
<b>35℃以上</b>	0.83	1.37	3.21	1.95	1.22	0.41	2.50	
95%信頼区間	0.2046 3.3407	0.5069 3.7048	1.3097 7.8448	0.7218 5.2860	0.3874 3.8221	0.0574 2.9383	1.0209 6.1025	
Exact P Asymptotic P	- 0.7894	- 0.5346	- 0.0108	- 0.1874	- 0.7368	- 0.3755	- 0.0449	
<b>36℃以上</b>	0.92	1.08	9.22	2.54	0.69	0.70	2.26	
95%信頼区間	0.2272 3.7121	0.2678 4.3792	3.3656 25.2474	0.8080 8.0122	0.0962 4.9311	0.0984 5.0436	0.8358 6.1267	
Exact P Asymptotic P	- 0.9049	- 0.9112	- 0.0000	- 0.1106	- 0.7105	- 0.7272	- 0.1081	

図 2 1 温暖時期の最高気温と糖尿病既往のない全年齢患者群におけるくも膜下出血発症との関連

(2) 今後複数の協力施設と連携してデータ解析を行うために「共通システム」を無償配布する  
目的で行う、学術的なシステムの公表・導入可能な施設への呼びかけ

本年度の研究実施計画にも記載あるように、本研究で開発した「共通システム」は、病院情報システムのデータと、環境省が保有する環境データ（環境省大気汚染物質広域監視システムのデータ等）の統合及び解析を可能にするシステムであり、今後、環境因子が疾病構造に与える影響を継続的にモニタリングし得る基盤となるシステムである。このため、本年度の成果として、当該システムを公開し、照会に応じて配付する日本語版マニュアル一式等を完成し、複数の協力施設と連携してデータ解析を行い得る環境の整備を図った。具体的な公開方法として、下記 2 通りの方法を用いた。

① 研究会における公表

まず、本研究班が主体となっている、電子カルテデータ・環境データを統合した疾病モニタリングに関する研究会主催の講演会において、参加者に「共通システム」を公表した（「医療分野におけるオープンソースソフトウェアについての現状と最近の問題点について、GNU Health の動向を含めて」主催；電子カルテデータ・環境データを統合した疾病モニタリング



に関する研究会，於 国立国際医療研究センター，講演：京都大学 医学研究科／EHR 共同研究講座／特定講師 小林慎治，平成 28 年 11 月 18 日）。

さらに広く、同様の電子カルテシステムを擁して、地域の中核的な医療センターとなっている、国立病院機構の病院群に「共通システム」をアピールし、学術的な公開を行う目的で、国立病院機構の査読有の学術誌である「医療」にシステムの概要とソフトウェア部分の無償配布、導入希望施設を広く募る旨の呼びかけを明記した原著論文を投稿し、現在印刷中である。下記にその内容の一部を紹介する。

② 医療分野におけるオープンソースソフトウェアについての現状および病院コホートを利用した医療情報と外部データ突合用データマイニングシステムの紹介

英語表題：Using FLOSS (Free/Libre/Open Source Software) for Medical Informatics and Data-Mining, Including Hospital Cohort, External Data, and Other Environmental Factors

著者名 星野 隆之	Takayuki Hoshino	1)
小林 慎治	Shinji Kobayashi	2)
植田 栄子	Teruko Ueda	3)
平山 雄一	Yuichi Hirayama	4)
松下 由実	Yumi Matsushita	1)

所属 1) 国立国際医療研究センター National Center for Global Health and Medicine  
 2) 京都大学 Kyoto University  
 3) 青森公立大学 Aomori Public University.  
 4) 株式会社システム情報パートナー System Information Partner, Inc.

要旨

情報通信技術の発達に伴い、医療情報の電子化と集積が急速に進み、電子カルテには日々の診療や保険請求業務を通じて膨大な医療情報が蓄積されている。これらの利用は電子処方箋や画像データの関連医療施設同士での共有、患者による診療履歴の閲覧等の形で還元されはじめている。さらに、これらの基盤を集約し、個々人の生涯にわたる健康記録を個人情報に配慮した形で抽出、二次的に活用する EHR (Electronic Health Record) の構築も進められつつある。EHR に集約された情報はビッグデータとして臨床疫学的に個々の中核医療機関が主体的に探求することができれば、医学的知見を新たに還元することに役立つと考えられる。本稿では、まず、EHR 構築において注目されている医療分野におけるオープンソースソフトウェアの現状と問題点について概説し、次いで、国立国際医療研究センターで開発、稼働しはじめた、病院コホートを利用した医療情報と外部データ突合用データマイニングシステムについて紹介する。

キーワード (和) 電子カルテ、オープンソースソフトウェア、臨床疫学

キーワード (英) Electronic health record, FLOSS (Free/Libre/Open Source Software), Clinical epidemiology

## はじめに

情報通信技術の発達に伴い、医療情報の電子化が急速に進み、電子カルテには日々の診療や保険請求業務を通じて膨大な医療情報が蓄積されつつあり、電子化されたデータを国家規模で収集する EHR(Electronic Health Record)の構築も試みられつつある。さらにこの EHR に蓄積された医療ビッグデータを個人情報に配慮した形で抽出し、臨床疫学的な探求を個々の中核医療機関までもが主体的に行うことができれば、医学的な知見をさらに臨床へ還元することに役立つと考えられる。本稿では、まず、EHR 構築において注目されているオープンソースソフトウェア(FLOSS: Free/Libre/Open Source Software)について概説し、医療分野における FLOSS の応用として日本の ORCA Project、アジアにおける FLOSS、中でも近年ジャマイカ、ラオス、フィリピンで戦略的に導入されつつある GNU Health について概説する。また、今後の FLOSS とその問題点についても最新の知見を交えて説明する。次いで、医療ビッグデータの国立病院機構などの中核病院での二次利用や相互連携を視野に、国立国際医療研究センターで開発、稼働しはじめた、病院コホートを利用した医療情報と外部データ突合用データマイニングシステムと、ソフトウェアの無償提供について紹介する。

## 医療分野における FLOSS について

本項では、まず FLOSS とはなにかについて概説し、その医療分野における応用と最近の問題点について本邦および諸外国の動向を含め概説する。主な内容は下記。

### ア FLOSS とは

#### イ 医療分野での FLOSS について

FLOSS に期待されること

(ア)開発コストの低減

(イ)標準化の推進

代表的な医療分野での FLOSS

- ・ OpenVistA : アメリカ退役軍人病院で開発された電子カルテシステム
- ・ GNU Health : Tryton(ERP)をベースとして開発された電子カルテシステム
- ・ OpenMRS : HIV, Tbc の管理のために開発された低開発国向け電子カルテシステム
- ・ DHIS2 : 疾病管理レポジトリ

#### ウ 日本における医療分野での FLOSS 活用

日本で開発された医療 FLOSS の主なもの

- ・ OpenDolphin : 地域医療連携システムのクライアントとして開発され、汎用電子カルテシステムとして独立している
- ・ NOAH : 大橋産婦人科で作成された PHP ベースの電子カルテ
- ・ tkDerm : 犬塚医師による皮膚科専門画像データベース

#### エ アジアにおける医療分野の FLOSS 実例

- ・ DHIS2 : 国境なき医師団で統一的に利用されている患者登録レポジトリであり、患者個人の疾患登録をベースに地域での疾病状況などの統計データ地図データを図示することができる。
- ・ OpenHIE : IHE 標準の FLOSS 実装で、病院間でのデータ連携を要因するソフト

ウェアである。

- GNU Health

近年ジャマイカでの国家戦略としての導入のほか、ラオス、フィリピンの病院施設等 350 ヶ所で採用されている GNU Health は、電子カルテのみならず、医療および福祉の情報を包括的に扱い、疫学的な解析も視野に入れたシステムとなっている。GNU Health は、Luis Falcon により 2008 年より開発開始され、Tryton という ERP (Enterprise Resource Planning) パッケージをベースとしている。病院単体あるいは複数の病院を連携して利用可能であり、個人および集団の健康管理、患者属性情報、家族単位、地域ごとなど、患者管理として、社会的経済的状況、生活歴、外来診察、入院、検査だけでなく、保健所業務として、予算、資産、検査施設、病院などの管理が可能であり、情報管理では、報告書作成、人口動態や疫学統計が可能である。GNU Health 日本語翻訳は、ボランティアで運営されており、一度は 100% 翻訳された時期があるが、このたび GNU Health 3.0 にアップデートされている。今後、多くの施設で利用される場合は、翻訳の監修などの作業が必要になることが考えられる。GNU Health 3.0 では、さらに、スマートフォンや Web からアクセス可能になるなど、新規機能が追加された。

#### オ 医療分野での FLOSS 活用の問題点

病院情報システムからの臨床データと外部データの突合解析が可能なデータマイニング用「共通システム」開発とこれを用いた疾病構造に関わる「環境モニタリングシステム」の紹介

前項で紹介されたように、世界的に医療情報の二次利用を目指し、医療情報をデータマイニングするソフトウェアが求められ、FLOSS の形で提供されつつある。一方で、医療 FLOSS が抱える諸問題、とくに医薬品医療機器等法の適用や運用・改定コスト、ソフトウェアの医療分野での利用や責任も大きな問題である。

このたび、われわれ国立国際医療研究センターでは、環境省地球環境保全等試験研究事業の、「環境因子の変化に伴う疾病構造変化モニタリングと中長期環境モニタリングおよび暴露調査結果を用いた環境がヒトへ与える影響の解析を行う病院コホートを利用したデータマイニングシステムの研究」により、病院情報システムからの臨床データと外部データの突合解析が可能なデータマイニング用「共通システム」を開発した。このシステムを用いた、電子カルテビッグデータと環境ビッグデータ突合解析による、疾病構造に関わる「環境モニタリングシステム」の具体例を図 1 (I-2.-図 1 参照) に示す。

われわれが開発した「共通システム」では、病院電子カルテやレセプトデータなどの病院情報ビッグデータからから情報抽出する際、抽出と同時に連結不可能匿名化を行い、突合用臨床データベースに再構築するため、個人情報非安全エリアに持ち出さずに、疫学研究が可能となる。国立国際医療研究センターでは、病院電子カルテシステム「Mega-Oak™ NEC Corporation, Tokyo, Japan」から臨床データを抽出している。Mega-Oak などの大病院向け電子カルテシステムの多くは電子カルテメーカーが独自に開発し、商品化したものである。このため、データベース化して外部データ、本研究の具体例では外部の公開環境データとの突合を行うためには、データ形式の標準化処理が必要となる。今回われわれは医療情報の「標準化ストレージ」という概念に着目し、すべての医療機関を対象とした医療情報の交換・共有による医療の質の向上を目的とした「厚生労働省電子的診療情報交換推進事業」に準拠し、

標準化の方法として、SS-MIX (SS-MIX : Standardized Structured Medical Information eXchange software version 2 , Consortium for SS-MIX Dissemination and Promotion, Kanagawa, Japan)を用いてデータベースの作成を行った。本臨床データベースから解析用に条件づけて抽出可能な臨床データ項目は、全ケースでの年齢、性別、郡市までの居所、保険種別、初受診日、初診日、再診日、受診診療科、病名、発症日、発症日を含む前7日移動平均データ、病名終了日、転帰(治癒、治療継続、中断、死亡)、一部入力があったケースの診療計測データ(収縮期血圧、拡張期血圧、脈拍、体重など)、検査項目、検査結果、検査日時、投薬内容、投薬日時、処置内容、処置日時などである。例えば、このシステムで抽出した臨床データと、環境データの公開されたデータソースとして、環境省大気汚染物質広域監視システム「そらまめ君」(<http://soramame.taiki.go.jp/> 環境省)、環境省花粉観測システム「はなこさん」(<http://kafun.taiki.go.jp/> 環境省)および気象庁気象観測データ(<http://www.jma.go.jp/jma/menu/menureport.html> 気象庁)を利用し、臨床データと環境データの突合解析を行うことが可能であり、われわれ研究班でも、実際に本システムを稼働させて、既存研究で着目されつつある、大気中のPM2.5濃度と虚血発作の発現との関連性につき、電子カルテから抽出・構築した臨床データベースと、環境データベースを突合解析し、得られた知見を公表している<sup>(1)</sup>。

われわれが開発したシステムは、ソフトウェアとしての特許取得は目指さず、希望される他施設への提供および、今後の疾病構造と環境因子などの外部データとの突合解析研究の基盤となることを目指している。医療分野のソフトウェアの宿命として、各導入施設において実装後、さらに改善点をあげて改良・カスタマイズが不可欠であり、いわば使用しながら徐々に改善していくソフトウェア開発手法である、スパイラルメソッドの一過程というべきものである。本ソフトウェアをこの「医療」にご紹介する意義として、類似の病院情報システムを稼働させ、せつかく、病院コホート研究が可能な臨床データをかかえた国立病院機構の医療情報ご担当に情報提供する目的がある。臨床ビッグデータを用いた解析を行う際、症例数も多く集約され、地域に根差しており、高度な医療と医療情報システム、その担当者を要したわれわれ国立病院機構の医療機関は、ちょうどよいセンターと成り得る。その際、臨床疫学的な解析には、今後本ソフトウェアのような、電子カルテシステムと、外部データをつなぐ、突合解析ソフトウェアが不可欠と考える。

おわりに

本報告では、まず、医療分野におけるオープンソースソフトウェアの現状と問題点について概説した。次いで、国立国際医療研究センター病院にて開発、稼働しはじめた、病院コホートを利用した医療情報と外部データ突合用データマイニングシステムについて紹介した。本システムは、環境省地球環境保全等試験研究事業の、「環境因子の変化に伴う疾病構造変化モニタリングと中長期環境モニタリングおよび暴露調査結果を用いた環境がヒトへ与える影響の解析を行う病院コホートを利用したデータマイニングシステムの研究」により研究開発されたものであり、その実施要項に基づき、作成した共通システムを公開、頒布用情報提供することで、将来、協力施設とも連携してデータ解析を行い得るシステム環境が整い、温暖化等の地球規模の環境因子の変動が健康におよぼす影響をはじめ、外部因子が疾病構造に与える影響を臨床疫学の視点で明らかにす

るための、一手段となることを目指している。このため、われわれが開発し得たソフトウェアの範囲内において、日本語版マニュアル一式とともに、ソフトウェアの無償での提供を行っている。実装には個々の施設での医療情報担当者によるカスタマイズや施設個々の電子カルテとの接続作業が必要で、この際にはコストが発生する可能性が考えられる。提供希望施設のお問い合わせをお待ちする。(著者連絡先、別冊請求先明記)

### 本研究課題終了後の発展的展望

本研究課題を通じて創出した「共通システム」を運用継続し、環境因子として、地上観測の環境因子データにとどまらず、宇宙航空研究開発機構 (JAXA) と共働し、人工衛星によるリモートセンシングで得られた環境データも活用することで、これまでにない空間的整合性で、臨床データと環境データのマッチアップデータセットを構築し、新たな生気象学的知見を得ることが可能となる。本研究班の呼びかけに応じて、国立病院機構など、本センター同様の施設への「共通システム」導入や共同での解析が進むことで、医学的な知見を得るのみならず、将来的により精度の高い、健康に関する環境モニタリングのインフラが構築可能となる。

### 6. 引用文献

- (1) Hoshino T, Hoshino A, Nishino J. Assessment of associations between ischaemic attacks in patients with type 2 diabetes mellitus and air concentrations of particulate matter <2.5  $\mu\text{m}$ . *Journal of International Medical Research* 2016;44(3),639-655
- (2) Hoshino T, Hoshino A, Nishino J. Relationship between Environment Factors and the Number of Outpatient Visits at a Clinic for Nonallergic rhinitis in Japan, Extracted from Electronic Medical Records. *Eur J Med Res.* 2015 Jul 8;20:60

[研究成果の発表状況]

#### (1) 誌上発表 (学術誌)

- ① 星野 (隆), 鈴木, 古場, 平山, 星野 (綾), 南川, 松岡, 菊池, 新保, 清水: 医療 1 号, 2014, 「電子化された医療データの共有と二次利用の国際動向」
- ② Hoshino T, Hoshino A, Nishino J. Relationship between Environment Factors and the Number of Outpatient Visits at a Clinic for Nonallergic rhinitis in Japan, Extracted from Electronic Medical Records. *Eur J Med Res.* 2015 Jul 8;20:60
- ③ Hoshino T, Hoshino A, Nishino J. Assessment of associations between ischaemic attacks in patients with type 2 diabetes mellitus and air concentrations of particulate matter <2.5  $\mu\text{m}$ . *Journal of International Medical Research* 2016;44(3),639-655
- ④ 星野、小林、植田、平山、松下: 医療, 2017, (In Press) 「医療分野におけるオープンソースソフトウェアについての現状および病院コホートを利用した医療情報と外部データ突合用データマイニングシステムの紹介」
- ⑤ 五嶋研人; 金岡慧; 木谷将人; 星野隆之, *Medi: 情報処理学会論文誌: データベース (TOD)*, 75 号, 2017, (In Press, 雇用ポスドクによる論文発表支援) 「電子カルテ蓄積データと環境データの統合及びマイニング手法」

## (2) 口頭発表

- ⑥ 星野（隆）、松原、植田、平山、星野（綾）、南川：第 41 回日本行動計量学会、2013 年 9 月 3 日～6 日  
「二次データ利用による喘息症状と大気汚染物質の関連性～二次データ利用によるローコスト大規模調査を目指して～」
- ⑦ 星野（隆）、松原、植田、平山、星野（綾）、南川：第 54 回 大気環境学会年会、2013 年 9 月 18 日～20 日  
「二次データ利用による喘息症状と大気汚染物質の関連性の推認」
- ⑧ 星野（隆）、松原、植田、平山、星野（綾）：日本行動計量学会第 43 回大会、2015 年「電子カルテデータと大気汚染物質広域監視システム「そらまめ君」を利用した諸疾患・大気汚染関係のデータスクリーニングー上下四分位法の活用ー」
- ⑨ Takayuki Hoshino, Nozomu Matsubara, Teruko Ueda, Yuuichi Hirayama, Ayami Hoshino, Kazuo Minamikawa : MIE2014、HALIC CONGRESS CENTER Istanbul, TURKEY, 2014/8/31-9/3  
“Monitoring System of Disease Structure by Integrating HIS and 'AEROS' in Japan”
- ⑩ Takayuki Hoshino, Nozomu Matsubara, Teruko Ueda, Yuuichi Hirayama, Ayami Hoshino : MEDINFO 2015:eHealth - enabled Health,San Paulo,Brazil,2015 /8/19-23  
"Quartile" screening method to analyze the relationship between HIS and 'AEROS' in Japan”

## (3) 出願特許

なし

## (4) 受賞等

なし

## (5) 一般への公表・報道等

なし

## (6) その他成果の普及、政策的な寄与・貢献について

- ① 第 1 回関東医療情報勉強会，主催：日本医療情報学会関東支部会、関東医療情報技師会，2014 年 12 月 13 日
- ② 「医療分野におけるオープンソースソフトウェアについての現状と最近の問題点について、GNU Health の動向を含めて」主催；電子カルテデータ・環境データを統合した疾病モニタリングに関する研究会，於 国立国際医療研究センター，講演：京都大学 医学研究科／EHR 共同研究講座／特定講師 小林慎治にて「共通システム」を公表，2016 年 11 月 18 日