

## 全国健康保険協会の各支部における分析用データベース構築の必要性

東京支部 企画総務グループ長 田島 哲也

保健グループ 岡本 康子、尾川 朋子、

企画総務グループ 吉川 彰一、馬場 武彦

国際医療福祉大学大学院 准教授 小川 俊夫

奈良県立医科大学 教授 今村 知明

---

### 概要

保険者においてデータヘルス計画を効果的に遂行する為には、各保険者が保有する加入者の健診結果やレセプト等の各種データを組み合わせ、分析用データベースを構築することが望ましい。この分析用データベースの構築にあたり、個人情報漏洩リスクを最小限に抑えつつ、構築した分析用データベースを重症化予防事業の対象者の抽出など多角的に利用するためには、分析用データベースは保険者内で構築・運用すべきであると考えられる。

本報告は、東京支部が外部有識者の協力を得て、データヘルス計画に先駆け、平成23年度より行ってきた分析用データベースの構築と活用について概説する。分析用データベースの構築には、協会本部において作成され各支部に毎年提供されている

「健診受診者リスト」を用いた。東京支部に提供された平成21年度から24年度までの「健診受診者リスト」を統計解析ソフトSPSS（IBM社）により個人毎に名寄せし、分析用データベースを構築した。構築した分析用データベースを用いて、健診受診者の健康状態および医療費を分析した結果、健康状態や平均医療費が、特定保健指導の階層毎に異なることに加え、性・年齢階級によっても違いが見られることが明らかになった。

分析用データベースを用いて特定保健指導対象者の指導への参加有無と翌年以降の健康状態を分析した結果、指導参加群は不参加群よりも腹囲などの指標で翌年以降に改善が見られ、指導の効果が示唆された。

また、分析用データベースを用いることで、慢性腎臓病（CKD）の疑いのある加入者を的確に抽出し、介入することが可能であった。

分析用データベースはこのように即時性にも優れており、自治体からのニーズに応じた集計データの作成と提供においても、的確に対応可能であった。

さらに、分析用データベースは拡張性にも優れており、今後のデータ追加や新たな項目の追加など、様々な用途に応用可能だと思われる。

分析用データベースは、データヘルス計画の中核となるばかりか、今後の支部運営に必要な不可欠な情報となる可能性があり、今後より一層のデータの充実と活用が望まれる。

---

## 【背景と目的】

保険者は、レセプトをはじめとした加入者の健康に関する様々なデータを保有しているが、これらの豊富なデータが十分に活用されてきたとは言いがたいのが現状である。このような状況において、「日本再興戦略」（平成 25 年 6 月 14 日閣議決定）において、「全ての健康保険組合に対し、レセプト等のデータの分析、それに基づく加入者の健康保持増進のための事業計画として「データヘルス計画」の作成・公表、事業実施、評価等の取組を求める」ことが掲げられた。すなわち、保険者は保有しているレセプトや健康診査など各種のデータを組み合わせて分析し、その結果を用いた保健事業をはじめとした保険者の各種事業の実施に役立てることとなった。

全国健康保険協会東京支部（以下、東京支部）では、データヘルス計画に先駆け、平成 23 年度より有識者の協力を得て本部より各支部に提供されている「健診受診者リスト」の活用について検討を開始し、平成 24 年度には分析用データベースの構築を行い、その活用を開始している。この東京支部の経験と成果は協会けんぽの各支部のみならず、協会けんぽ全体のデータ活用に有益であると考えられる。

本報告は、東京支部で実施した分析用データベース構築の経緯と手法について概観したうえで、分析用データベースの活用事例についても概説する。また、今後の協会けんぽ全体及び各支部にとって必要な分析用データと資源、分析用データベースのさらなる活用方法について考察を実施する。

## 【方法】

本報告においては、まず分析用データベース構築に向けて行った様々な検討を概観したうえで、分析用データベース構築の方法について概説する。次に、構築された分析用データベースを用いた分析と活用の事例として、特定保健指導の効果分析と CKD 重症化予防への活用、自治体への情報提供について紹介する。これらを踏まえて、今後の協会けんぽ全体及び各支部にとって必要な分析用データと資源、分析用データベースのさらなる活用方法について考察を実施する。

分析用データベース構築に向けた各種検討は、平成 23 年度に有識者の協力のもとに行われた。この検討においては、支部内の資源を活用することを目的とし、支部で利用可能なデータを精査し、また支部で利用可能な PC やソフトウェアについても検討した。本報告では、これらの検討結果を概説する。分析用データベース構築については、分析用データベース構築の手順書として協会内で共有された手法について概説する。特に、我々が用いた手法の特徴について、他の方法と対比しつつ要約する。分析用データベースの活用事例については、各学会での発表や原著論文として掲載された成果を中心に概説する。

## 【結果】

### (1) 分析用データベース構築に向けた検討

分析用データベース構築にあたり、分析用データベース構築に用いるデータの確保、構築主体の策定と構築に必要な資源の確保などについて検討を実施した。

#### i) データの確保：

平成 23 年度に有識者の協力を得て、東京支部で利用可能な各種データと資源について検討を開始した。分析に用いるデータは、生活習慣病予防健診と特定保健指導の結果及びレセプトとし、「健診受診者リスト」を用いることとした。この「健診受診者リスト」は、平成 21 年度より本部から各支部に csv ファイルとして毎年提供されており、各年度提供の「健診受診者リスト」には、健診及び指導については 12 ヶ月分、レセプトについては、18 ヶ月分の健診・指導・レセプト情報が名寄せされて格納されている。「健診受診者リスト」は提供された状態でも分析可能であるが、単年度データであるため、そのままでは経年的な分析は不可能である。そのため、経年分析を行うためには複数年度の「健診受診者リスト」を名寄せ・突合する必要があることが明らかになった。

#### ii) 分析用データベースの構築主体：

分析用データベース構築の方法について、外部有識者や民間事業者を利用する外注方式と協会けんぽで実施する内製方式について検討した。分析用データベース構築と活用の際には、分析結果の即時活用の観点から、健診や指導、レセプトデータに含まれる個人情報等を全て保有したままで構築・活用することが重要であり、また必要に応じてデータの追加や修正が支部内で可能なことが重要である。したがって、個人情報の漏洩リスクを最小限に抑えつつ、重症化予防事業の対象者の抽出などに多角的に利用するためには、分析用データベースは有識者のサポートのもとで保険者内において保険者自身により構築されることが適切と判断した。

#### iii) データベースの構築と活用に必要な資源の確保：

分析用データベース構築と活用を支部内で実施するためには、データベース構築と分析にかかる担当者、データベース構築と分析用の PC およびソフトウェアの確保が必須である。分析担当者は、通常業務を行いつつ実施可能な範囲で作業を実施する必要があるため、そのためデータベース構築や分析の省力化が必須と考えられた。また、分析のための PC やソフトウェアの追加導入は様々な事情から困難と判断し、通常業務に使用している PC と既存ソフトウェアを利用する方向で検討した。

なお、東京支部に提供される「健診受診者リスト」には、毎年約 60 万人分の受診

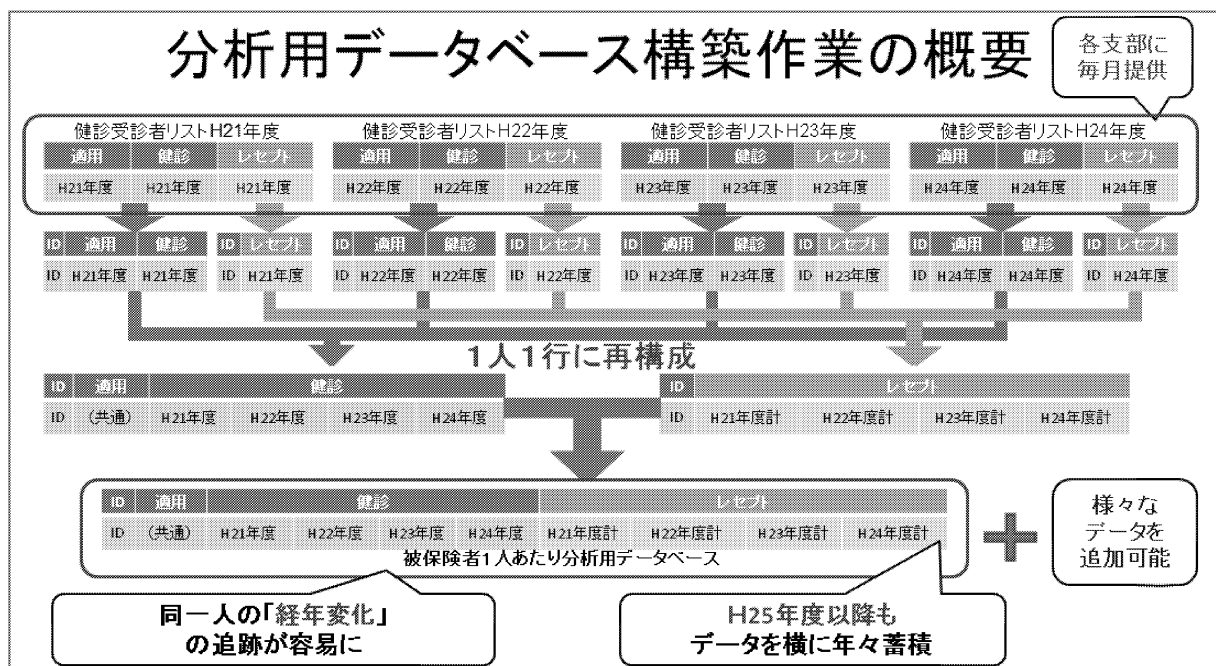
者情報が含まれている。各年度のデータ容量は約 2GB であり、健診受診者数の増加に伴い年度あたりに収載された人数・データ容量は年々増加する傾向にある。そのため、本部から提供される「健診受診者リスト」を Microsoft Excel や Access で処理して複数年度のデータを格納した分析用データベースを構築することは、東京支部の場合には困難であるが、各支部に 1 ライセンスずつ提供されている統計解析ソフトウェア SPSS (IBM 社) を用いることで、大容量データの処理と分析が可能になることが明らかになった。また、SPSS のシンタックスと呼ばれるプログラムを活用することで各種処理の一部を自動化することが可能であり、その結果として分析用データベースの構築と分析実施における作業の省力化も可能であると判断された。

以上より、東京支部の通常業務に使用している PC に本部より提供された SPSS とデータを格納し、有識者の協力のもとに東京支部の職員が分析用データベース構築と活用を行うことになった。

## (2) 分析用データベース構築の手順

東京支部では、平成 24 年度に「健診受診者リスト」の平成 21 年度から 23 年度までの 3 カ年分のデータを、SPSS のシンタックスを用いて個人毎に名寄せし、分析用データベースを構築した。その際に、年度間で重複するデータを処理したほか、レセプトデータについては年度あるいは年月単位で集計した。なお、名寄せに使用する結合キーには、健康保険被保険者証の協会支部番号 (保険者番号) ・事業所記号・被保険者番号・被扶養者番号から作成した独自の個人 ID を用いた。平成 25 年度には、分析用データベースに「健診受診者リスト」の平成 24 年度のデータを付加し、4 カ年分のデータを格納したデータベースを構築した。また、生活習慣病予防健診の非受診者のレセプト情報も含む全レセプト情報を分析用データベースに付加し、医療費分析を可能とした。なお、これらデータベース構築の手順や注意点、SPSS で使用するプログラムなどについては「SPSS を用いた健診受診者リストの分析手順書」として取りまとめ、協会内部で共有した。

(図 1：分析用データベース構築作業の概要)



### (3) 分析用データベースの活用事例

構築した分析用データベースの最大の特長は、名寄せによって同一人物のデータが横一行に並べられている為、その経年変化を容易に分析できることである。この分析用データベースを用い、加入者の健康状態や医療費の特性を把握したほか、経年変化などの各種分析を実施した。更に、重症化予防などの保健事業の対象者抽出や、提携する地方自治体からのニーズに応じた集計データの作成・提供などにも活用した。以下に分析用データベースを活用した主な成果について概説する。

(図 2 : 分析用データベース画面イメージ) 横 1 行に同一人のデータが並ぶ。

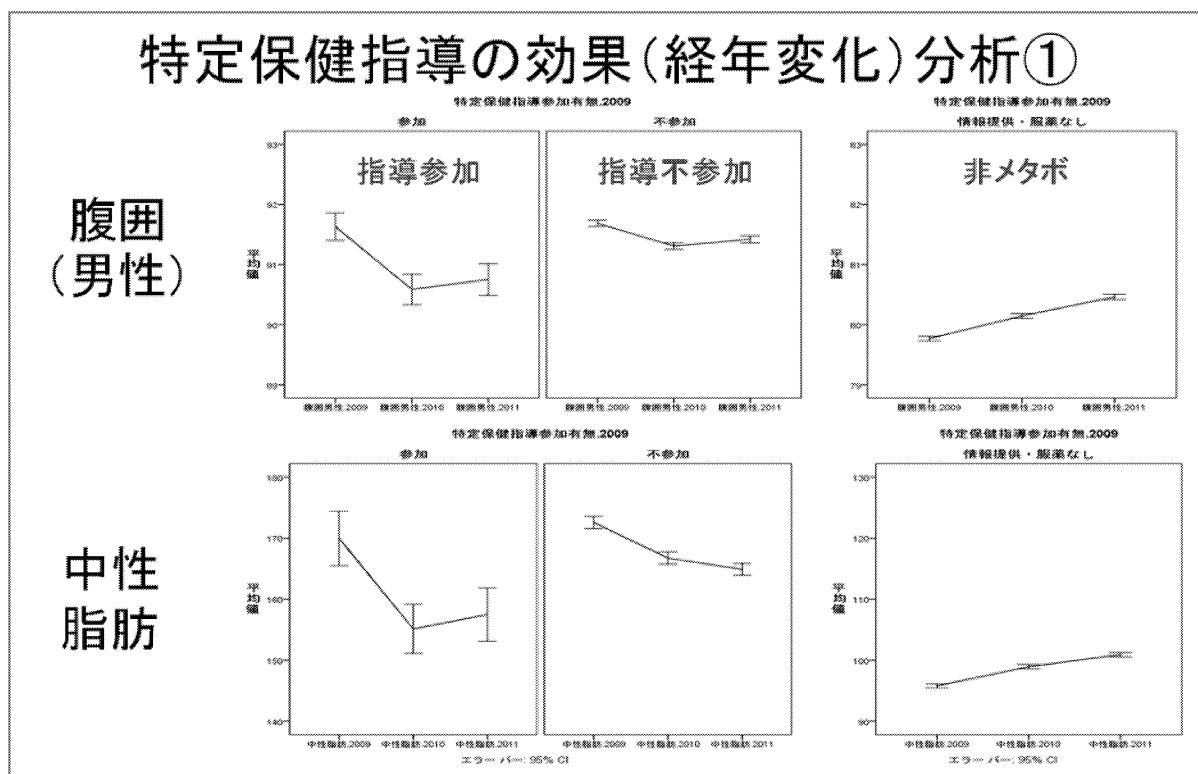
	最高血圧 2009	最高血圧 2010	最高血圧 2011	最高血圧 2012	最高血圧階級 2009	最高血圧階級 2010	最高血圧階級 2011	最高血圧階級 2012
48690	116.00	147.00	118.00	123.00	60~119mmHg	140~159mmHg	60~119mmHg	120~129
48691	106.00	126.00	135.00	119.00	60~119mmHg	120~129mmHg	130~139mmHg	60~119
48692	106.00	106.00	98.00	100.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48693	130.00				130~139mmHg			
48694	100.00	100.00	110.00	110.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48695	144.00	128.00	130.00		140~159mmHg	120~129mmHg	130~139mmHg	
48696	112.00	104.00	114.00	108.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48697	85.00	92.00	90.00	86.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48698	94.00				60~119mmHg			
48699	111.00	96.00	92.00	96.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48700		116.00	118.00			60~119mmHg	60~119mmHg	
48701	106.00				60~119mmHg			
48702			102.00	110.00			60~119mmHg	60~119
48703	118.00	110.00	121.00	116.00	60~119mmHg	60~119mmHg	120~129mmHg	60~119
48704	103.00	106.00	105.00	100.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48705	140.00	130.00	126.00		140~159mmHg	130~139mmHg	120~129mmHg	
48706	112.00	104.00	108.00	96.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48707	97.00	104.00	103.00	101.00	60~119mmHg	60~119mmHg	60~119mmHg	60~119
48708	138.00	142.00			130~139mmHg	140~159mmHg		
48709	133.00				130~139mmHg			
48710	116.00	120.00	118.00	120.00	60~119mmHg	120~129mmHg	60~119mmHg	120~129
48711			146.00				140~159mmHg	
48712	104.00	112.00	114.00		60~119mmHg	60~119mmHg	60~119mmHg	

### i) 特定保健指導の効果分析

分析用データベースから平成 21 年度の特定保健指導の対象者を抽出し、中断者を含む指導参加群と不参加群に区分し、対照群として情報提供と判定された人のうち服薬のない群を抽出した。これら分析対象群の平成 21 年度から 23 年度の各健診項目の判定結果の変化を比較した。また、同期間の健診数値の平均値の経年変化を分析した。

各健診項目の判定結果の経年変化については、男性の腹囲・BMI など、指導参加群の改善傾向が不参加群よりも有意に高かった ( $p < 0.01$ )。また、健診数値の経年変化については、男性の腹囲・中性脂肪などで指導参加群の改善傾向が不参加群よりも有意に高かった ( $p < 0.01$ )。この傾向は、積極的支援群と動機付け支援群に細分化してもほぼ同様であった。一方、対照群ではほぼ全ての健診項目で悪化傾向であり、これらより特定保健指導への参加が健康状態を改善した可能性が示唆された。なお、本分析の概要は第 72 回日本公衆衛生学会総会にて報告したほか、「厚生指標」2014 年 1 月号に論文として発表した<sup>1)</sup>。

(図3：特定保健指導後の経年変化)



### 特定保健指導の効果(経年変化)分析②

積極的支援

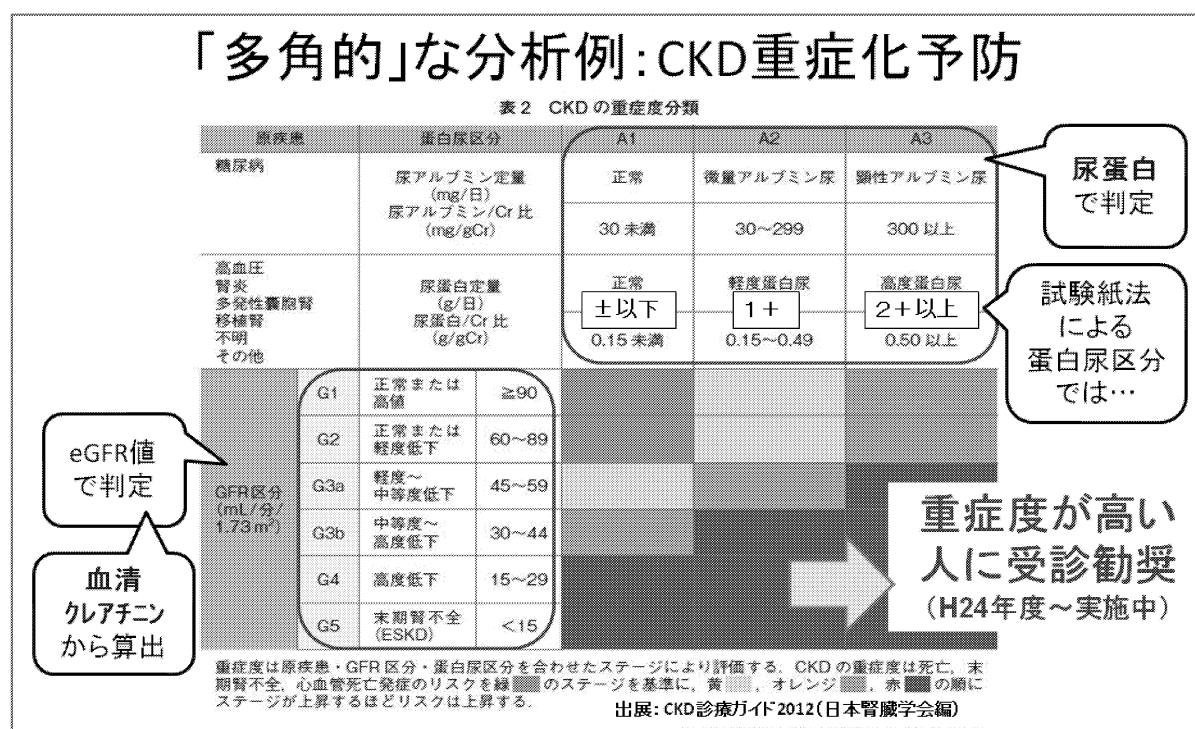
	2009・2010両年度受診者(人)				2009・2011両年度受診者(人)				改善割合(%)		悪化割合(%)	
	総数	改善	変化なし	悪化	総数	改善	変化なし	悪化	2009・2010年度	2009・2011年度	2009・2010年度	2009・2011年度
腹囲(男性)												
参加	2 358	379	1 971	8	2 039	325	1 704	10	16.1	15.9	0.3	0.5
不参加	50 130	5 707	44 110	313	43 189	5 278	37 622	289	11.4	12.2	0.6	0.7
腹囲(女性)												
参加	133	25	108	2	121	25	95	1	18.8	20.7	1.5	0.8
不参加	3 363	581	2 696	86	2 888	503	2 301	84	17.3	17.4	2.6	2.9
BMI												
参加	2 489	201	2 204	84	2 160	191	1 864	105	8.1	8.8	3.4	4.9
不参加	53 458	2 948	48 186	2 374	46 057	3 197	40 243	2 617	5.5	6.9	4.4	5.7
最高血圧												
参加	2 491	395	1 842	254	2 161	353	1 569	239	15.9	16.3	10.2	11.1
不参加	53 508	8 076	39 907	5 525	46 084	7 287	33 430	5 367	15.1	15.8	10.3	11.6
最低血圧												
参加	2 491	375	1 900	216	2 161	346	1 599	216	15.1	16.0	8.7	10.0
不参加	53 504	8 005	39 801	5 698	46 084	7 290	33 280	5 508	15.0	15.8	10.5	12.0
中性脂肪												
参加	2 491	501	1 807	183	2 160	414	1 567	179	20.1	19.2	7.3	8.3
不参加	53 450	9 012	40 123	4 315	46 030	8 346	33 849	3 835	16.9	18.1	8.1	8.3
HDL												
参加	2 491	194	2 188	109	2 161	186	1 883	92	7.8	8.6	4.4	4.3
不参加	53 492	3 659	47 384	2 449	46 074	3 271	40 622	2 181	6.8	7.1	4.6	4.7
空腹時血糖												
参加	2 405	352	1 890	163	2 085	293	1 635	157	14.6	14.1	6.8	7.5
不参加	51 937	6 857	40 622	4 458	44 665	8 068	34 304	4 293	13.2	13.6	8.6	9.6
HbA1c												
参加	354	29	290	35	312	35	256	21	8.2	11.2	9.9	6.7
不参加	5 660	356	4 827	477	4 673	348	3 931	394	6.3	7.4	8.4	8.4
喫煙												
参加	2 490	147	2 312	31	2 160	195	1 935	30	5.9	9.0	1.2	1.4
不参加	53 430	2 926	49 875	629	46 048	3 958	41 450	640	5.5	8.6	1.2	1.4

ii) CKD 重症化予防への活用事例

慢性腎臓病（Chronic Kidney Disease: CKD）は重症化すると人工透析が必要になる場合があり、また心不全などの重大な心血管疾患の発症リスクが大幅に増大し、患者のQOL（生活の質）を大きく損ねる疾病と言われている。そのためCKDの早期発見・早期治療が重要であるが、CKDには自覚症状が乏しい為、健診などで早期発見する必要がある。

CKDには、日本腎臓学会が定めた重症度分類が存在するが、その判定にはGFR（糸球体濾過量）と尿蛋白の2つの検査結果が用いられている。このうちGFRは血清クレアチニンから算出したeGFR（推算糸球体濾過量）を用いることが可能である。eGFRは血清クレアチニン値・年齢・性別に基づき一定の計算式で算出される。血清クレアチニンは特定健康診査の検査項目には含まれていないが、協会の生活習慣病予防健診の検査項目には含まれている。尿蛋白は、試験紙法による定性検査で分類することが可能で、特定健康診査の検査項目に含まれている。

(図4：慢性腎臓病（CKD）の重症度分類表)



CKDの定義では、腎機能の低下が3ヶ月以上持続することが条件として挙げられていることから、分析用データベースを用いて健診を2年連続で受診した人で、これらの検査数値が2年連続で悪い人を、腎機能の低下した状態が3ヶ月以上持続している可能性が高くCKDが強く疑われる群、すなわちCKD疑い群として抽出した。

抽出したCKD疑い者に、過去2年間の実際の検査数値を記載した早期治療を勧奨す



る通知文書を送ったところ、通知対象者の7～8名につき1名程度が通知を機に治療を開始した。

(図5：慢性腎臓病（CKD）重症化予防事業の概要)

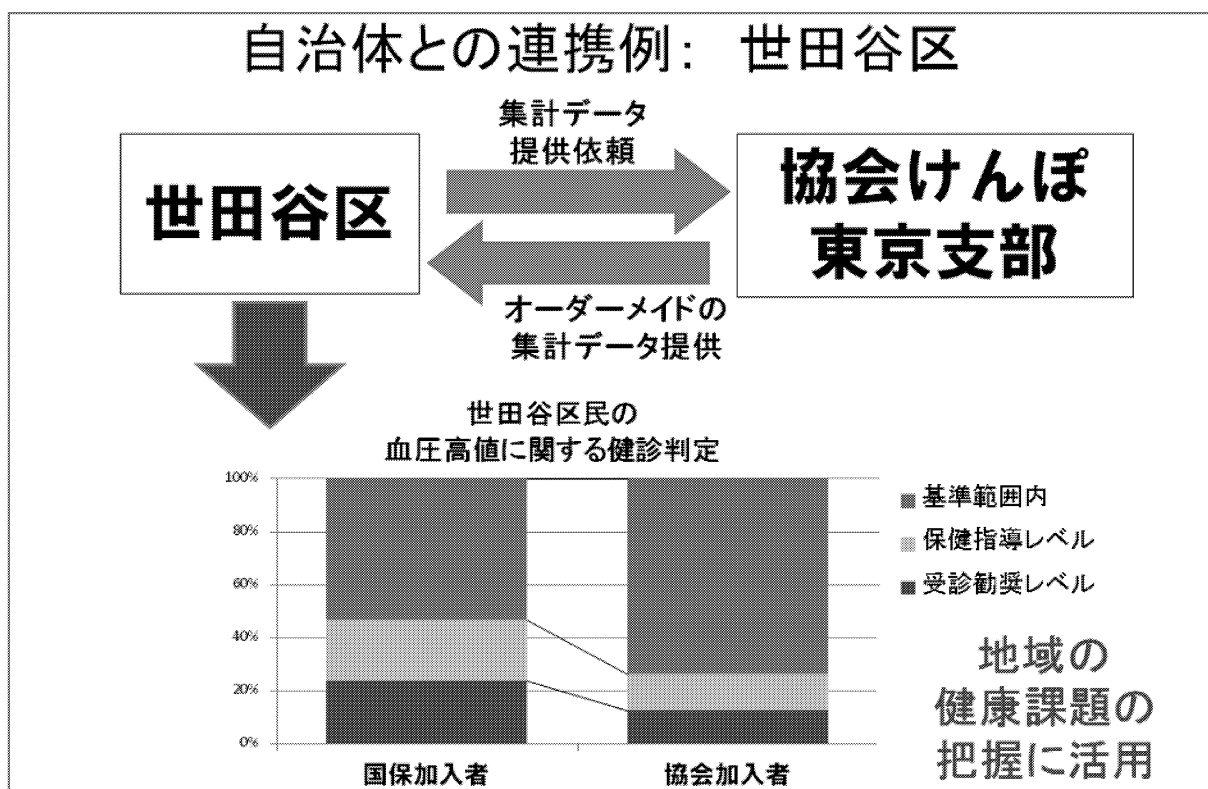
## CKD(慢性腎臓病)重症化予防事業 概要

- 目的：CKD重症化による人工透析導入・QOL低下の予防
- 方法：文書による早期受診勧奨
- 対象者：CKD重症度が2年連続で高い、人工透析未導入者  
平成24年度 3,466名  
平成25年度 5,255名（対象範囲を軽症者に拡大）
- 効果：アンケート調査により効果測定（H24年・H25年）
  - ・回答者数……………840名・842名
  - ・受診勧奨により新たに治療開始…105名・139名→成果
  - ・受診勧奨の前から既に治療中……487名・668名
  - ・協会保健師に健康相談を希望……114名・144名→実施
  - ・自覚症状がない為、治療せず※…116名・161名→要対策  
（※糖尿病、高血圧などCKD以外では治療をしている者を含む。）
- 課題：・アンケート調査未回答者の動向も把握する為、  
レセプト情報を活用した効果検証が必要  
・医療費への効果検証には長期間の追跡が必要

### iii) 地域との連携への活用

分析用データベースは、東京都や世田谷区など東京支部と提携する自治体からのニーズに応じた集計データの作成と提供においても活用されている。例えば、被保険者台帳に含まれる郵便番号情報から任意の自治体に居住する被保険者のみを抽出し、その健康状態や医療費を分析することも可能である。このように、協会けんぽの保有している加入者情報と、自治体が保有している国民健康保険の加入者情報を組み合わせることで、より多くの地域住民の健康状態の把握が可能となると考えられる。

(図6：自治体からのニーズに応じた集計データ提供例)



## 【考察】

平成 23 年度からの分析用データベース構築と活用に関する東京支部の取り組みにより、経年変化など多角的な分析が可能となる分析用データベースを、日常業務に利用している PC と SPSS を用いることで構築可能であることが明らかとなった。特に分析用データベースの構築にあたり、SPSS のシンタックスを用いることによりある程度の自動化が可能になったことに加え、Microsoft Excel や Access では通常、扱うことが困難な大容量のデータを扱うことが可能になった。さらに、その構築と分析の手順書を作成したことで、協会の各支部における分析用データベースの構築と活用が可能になった。仮に、協会けんぽの各支部が同じ手法で分析用データベースを構築し、その横断的な分析が可能になれば、より詳細かつ広範囲な分析が可能になると考えられる。

分析用データベースを構築することで、健診結果やレセプト単独の分析のみならず、健診結果や保健指導結果、レセプトを組み合わせた分析が可能になったうえ、複数年のデータを名寄せしたことから経年分析も可能となった。この結果として、分析用データベースを用いることで健診受診者の健康状態や医療費の正確な把握が可能になり、エビデンスに基づいた多角的なデータ解析が可能になった。

東京支部では、このような分析用データベースを用いて健診受診者の健康状態および医療費を分析した結果、健康状態や平均医療費が特定保健指導の階層毎に異なること、また性・年齢階級によっても違いが見られることを明らかにした。また、特定保健指導対象者の指導への参加有無と翌年以降の健康状態を分析した結果、指導参加群は不参加群よりも腹囲などの指標で翌年以降に改善が見られ、指導の効果が示唆された。

また、分析用データベースを支部内で構築・運用することで、個人情報漏洩のリスクを最小限に抑えることが可能であり、さらに個人情報を含む分析用データベースを活用することで、保健事業として介入すべき群をより正確かつ迅速に抽出することが可能になった。その結果として抽出した群に対してより効果的な介入が可能となり、保健事業の効果の向上に貢献できると考えられる。東京支部では CKD に着目し、分析用データベースを CKD 疑い群の抽出に活用し、抽出した群に対して介入を行うことで CKD の重症化予防に寄与できたと考えられる。

分析用データベースの特性の一つに、分析対象群の抽出も容易であることが挙げられる。例えば東京支部において、自治体からの要請により特定地域に居住している協会けんぽ加入者を特定したうえで健康状態や医療費について分析し、その結果を地方自治体と共有することで、当該地域住民の健康状態や医療費の状況をより正確に把握することが可能になったと考えられる。このような協会外部からの集計データ提供依頼などに対しても、分析用データベースを活用することで適時に対応可能となった。

分析用データベースは拡張性にも優れている。健診や指導、レセプトデータは毎月あるいは毎年追加されるため、分析用データベースの拡張性が重要となるが、東京支部が構築した SPSS を用いた分析用データベースでは、健診や保健指導、レセプトデ

ータの追加も容易であり、分析用データベースを用いて継続した分析が可能である。さらに、がん検診結果や業種情報などの追加情報の付加も容易であり、これらの付加情報を用いてより多角的な分析とその結果の活用が可能となる。しかしながら、分析に用いている現状の PC の能力ではいずれ PC 側の限界がくると考えられ、分析用データベースの構築と活用に関する PC について検討すべきである。

なお、分析用データベースの構築と分析には、以下の課題が存在する。第一に、分析用データベースの構築と運用には、東京支部で作成した手順書によりある程度は簡略化・自動化できたが、かなりの労力と経験が必要である。第二に、東京支部の様に加入者数の多い支部においては、分析用データベース構築と運用に用いている現状の PC の性能が充分ではなく、今後のデータの更なる蓄積に伴い、支部で日常業務に利用している PC では性能不足になると考えられるため、分析用データベースの構築と活用に関する PC について検討すべきである。第三に、分析用データベースの構築に用いたレセプト情報は、「健診受診者リスト」に記載されている合計点数・主傷病名などのみであり、これら主要項目以外のレセプトデータは含まれておらず、その活用には限界が存在する。第四に、分析用データベースを用いた分析結果の解釈には専門的な知識が必要であり、そのための知識や経験を職員が習得するか、あるいは外部有識者を活用する必要がある。第五に、分析用データベースは加入者一人あたり一行のデータにより構成されているが、レセプトに記載されている疾病情報はレセプト一枚あたり複数記載されていることから、その処理は現在の分析用データベースのデザインでは限界がある。そのため、疾病毎の詳細分析を行うことは、現在の分析用データベースでは困難であり、今後データマイニング用のソフトウェアの導入と分析手法などについて検討が必要である。

東京支部が構築した分析用データベースは、その実用性と拡張性の高さから、データヘルス計画の中核として位置づけることが可能であり、今後東京支部のみならず、協会けんぽの各支部への導入が期待される。また、本部が主導して分析用データベースの構築を行うことが可能であれば、協会けんぽ全体の分析や支部間の比較分析なども可能になると考えられる。また、分析用データベースの構築により、協会全体で用いるデータの質の維持と向上がより容易になり、また支部間の分析用データの整合性が取れるようになると考えられる。

最後に、分析用データベースを構築した調査研究事業に先立つ平成 22 年度のパイロット事業の段階からコーディネーター役として東京支部をご指導頂き、データベース構築の手法・手順書の作成、分析デザインや分析の実施、学会・論文発表などに多大なご助力を頂いた、奈良県立医科大学の今村知明教授、並びに国際医療福祉大学大学院の小川俊夫准教授に、この場をお借りして篤く御礼を申し上げます。

【参考文献】

- 1) 吉川彰一・小川俊夫・馬場武彦・南友樹・尾川朋子・田島哲也・山根明美・今村  
知明 (2014).  
特定健康診査・特定保健指導の効果分析－全国健康保険協会東京支部における特  
定健康診査受診者の健康状態の年次変化－ 厚生の指標, 61, 33-40