

# レセプト・健診データ分析により実現する 保健事業の人的リソース重点配分

(株)NTT データ 技術開発本部  
高橋 彰子, 中川 慶一郎, 関根 純

NTT データマネジメントサービス(株) シェアードサービス事業本部  
恒松 直幸

平成 20 年 4 月から始まった新しい健診制度「特定健康診査・特定保健指導」の影響で、生活習慣病予防の取り組みもすっかり周知となった感があるが、国民医療費が年々増加傾向にある公的医療保険分野においては、かねてより疾病予防により医療費を削減しようとする動きが見られた。今回の報告では、その先駆けとも言える平成 16 年度経済産業省予算によるモデル事業を紹介する。

モデル事業では、97 団体の健康保険組合（以下、健保組合）が参加する「保険者機能を推進する会」と協力し、8 団体の健保組合を中心に、生活習慣病予防を実現するような保健事業の企画と、この保健事業の実施に際して人的リソース（保健指導人材）の重点配分を実現するようなレセプト・健康診断データの分析を行っている。このような根拠に基づく保健事業は EBH（Evidence Based Healthcare）と呼ばれているが、今回は、この EBH を支える分析に焦点を絞って紹介する。

また、今回紹介するようなリソースの重点配分を実現する分析は、保健分野にとどまらず幅広い分野に適用可能であるが、このように分野を越えて共有することのできるいわば「分析の定石」とでも呼ぶべきものを弊社では「分析シナリオ類型」として整備し、(株)数理システムの Visual Mining Studio を用いてテンプレート化しているので、これも紹介する。

## 1 はじめに

生活習慣病の増加などにより国民医療費が年々増加傾向（平成 18 年度 33.1 兆円、過去 15 年間で 1.5 倍に増加[1]）にあるなか、健保組合をはじめとする保険者には、保健事業を通じて被保険者一人ひとりの疾病の予防と全体としての医療費削減に努める役割が期待されている。

保険者の医療費負担は、医療機関により毎月請求される患者ごとの診療報酬明細書（以下、レセプト）に基づき決定される。従来レセプトは紙形式で授受されてきたが、平成 18 年 4 月の厚生労働省による省令を受け、平成 23 年度以降一部の例外を除き全国の医療機関からの請求はすべてオンラインシステム上で行うことが義務付けられ[2]、医療事務の効率化とレセプト情報の有効活用が期待されている。

今回紹介する疾病管理プログラムは、以上のような背景の下、健保組合による根拠に基づく保健事業（EBH: Evidence Based Healthcare）を支援する目的で実施した平成 16 年度経済産業省予算によるモデル事業である。この保健事業とは、被保険者に対する個別面談の実施や運動

施設の提供など生活習慣病の予防を目的とした保健指導のことを指しているが、モデル事業ではレセプトおよび健康診断データ（以下、健診データ）という根拠情報に基づき、高い効果が見込まれる被保険者らを抽出し、優先的に保健指導の対象とすることで、保健指導人材という希少な人的リソースの有効活用を図った。

なお、このようなリソースの重点配分を実現する分析は、保健分野にとどまらず幅広い分野に適用可能なものであるが、今回は、この疾病管理プログラムの概要に加え、このように分野横断的に共有することのできる「分析の定石 = 分析シナリオ類型」についても併せて紹介する。

## 2 疾病管理プログラムの概要

疾病管理プログラムとは、保健事業の計画・実行・評価の PDS サイクルである(図 1 参照)。DO 段階にあたる保健事業の実行に対して、計画のための事前分析が PLAN 段階、評価のための事後分析が SEE 段階にあたる。

事前分析の目的は、3 つに分けることができる。第一の目的は、「重症化の原因疾病の特定」であ

る。生活習慣病の場合、複数の疾病を併発することによって重症化していくのが一般的である。いったん重症化してしまうと健保組合に出来ることは限られるので、重症化の発端となる原因疾病を特定し、被保険者がそもそもこの疾病にかからないよう保健指導という予防対策を打つ。

モデル事業では、被保険者一人ひとりの疾病履歴が記録されているレセプトデータに時系列アソシエーション分析[3]を適用し、生活習慣病の重症化過程をモデル化(重症化モデルの構築)した後、この過程の発端に位置づけられる原因疾病として、高血圧性疾患(以下、高血圧)と糖尿病を特定(原因疾病の特定)した。

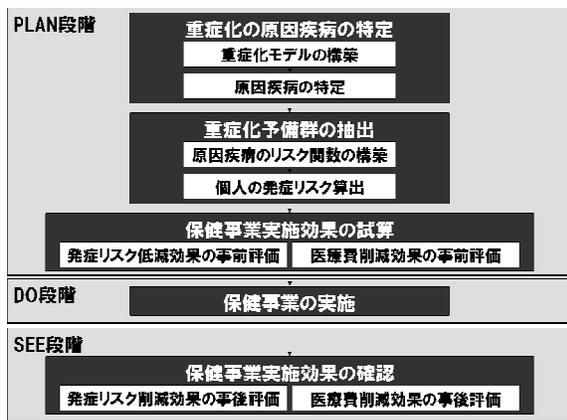


図1 疾病管理プログラムのPDSサイクル

事前分析の第二の目的は、「重症化予備群の抽出」である。予算の制約を考えると、すべての被保険者に対し一律に保健指導を実施することはできない。そこで、将来的に原因疾病を発症し、重症化の過程をたどることになるリスクの高い被保険者らを重症化予備群として抽出し、優先的に保健指導の対象とするのである。

ある人が将来生活習慣病を発症するリスクは、その人の遺伝的特性や生活習慣により異なるが、健診データはその影響を総合的に表したものと考えることができる。一方、レセプトデータはその人が生活習慣病を発症したか否かの結果を示してくれる。

モデル事業では、まず健診データとレセプトデータを突き合わせ、これにロジットモデル[4]を適用して原因疾病(高血圧および糖尿病)の発症リスクを予測する関数を構築(原因疾病のリスク関数の構築)した。その後、被保険者

一人ひとりの最新の健診データをリスク関数に投入することによって、高血圧および糖尿病の発症リスクを算出(個人の発症リスク算出)し、リスクの高い集団を保健指導の対象として抽出した。

事前分析の最後の目的は、「保健事業実施効果の試算」である。ここでは、前段階で抽出した集団に対して実際に保健指導を実施した場合に期待される効果をシミュレーションにより試算する。保健指導の効果としては、健診値改善による生活習慣病の発症リスク低減効果と、それに伴う将来の医療費削減効果が考えられる。

モデル事業では、たとえば糖尿病を発症するリスクの高い順に一定人数の被保険者に対して保健指導を行い、糖尿病と関連の強い健診項目の値を一定割合改善させたなどの仮定をおき、その場合に対象者一人ひとりのリスク値がどの程度低減されるかを試算(発症リスク低減効果の事前評価)すると同時に、全体として医療費がどの程度削減されるかについても試算(医療費削減効果の事前評価)した。

以上のようなシミュレーションを行うことで、事前に十分に効果を見極めたうえで保健事業の実施に移ることができる。つまり、効果の期待できる被保険者に対して重点的に保健指導人材を当てることができる。ここで、純粋に費用対効果を考えるならば、保健指導の対象者は発症リスク上位者ではなく、医療費削減効果の上位者ということになる。しかし、被保険者の健康を維持するのが健保組合の第一義であるということもあり、モデル事業では発症リスク上位者を保健指導の対象者とした。

最後に、保健事業実施後の事後分析であるが、これは、事前分析で試算した保健事業の効果を事後に測定・確認するものである。この部分については、モデル事業では机上検討しか行っていないため、今回は報告しない。

### 3 分析シナリオ類型の概要

以上では、レセプト・健診データ分析の事例を紹介してきたが、この事例で実施したようなリソースの重点配分を実現する分析は、幅広い分野に適用可能なものである。以下では、分野を越えて共有することのできる「分析の定石 = 分析シナリオ類型」の概要を紹介する。

いったん保健分野を離れて、金融分野に目を向けてみよう。金融ビッグバンにより規制緩和が進み、銀行が取り扱う商品はいまや投資信託や保険にまで広がった。このような業容の拡大は業態を超えた新たな競争を促し、その結果、銀行も個人顧客に対して積極的にダイレクト・マーケティングを展開するようになった。

顧客一人ひとりに最適な金融商品を薦めるということは、裏を返せば商品ごとに契約見込みの高いターゲット顧客を絞り込んでいるということである。このとき、人のライフステージを例えば「就職 結婚 子ども誕生 住宅購入 子ども進学」といった段階的なモデルで捉えると、顧客の取引履歴や属性、金融商品に対する価値観などのデータを用いてライフステージ毎にターゲット顧客を絞り込むことが可能となる。

以上で紹介した保健分野と金融分野の事例を比較すると、

- 重点的にアプローチすべきターゲットを見つけ出すという分析目的
- 疾病・取引の発生履歴や対象者の属性といったデータ形式
- 対象者の状態を段階的なステージにより捉える分析モデル

といった内容が類似している(図2参照)。当然そこでは同じような分析手法を用いて同じような意思決定を行っており、いわば「分析の定石」とでも呼べる分析パターンが存在する。



図2 保健分野と金融分野の事例比較

この分野を超えて共有することのできる分析パターンを我々は「分析シナリオ類型」と呼び、過去の分析事例に基づき8パターン(サブ類型も含めると12パターン)に分類しているの、表1にその概要を、表2に典型的な事例を示す。

表1 分析シナリオ類型の概要

分析シナリオ類型	概要	
① 予兆発見型	大量のデータから有るかの増進変化や状態変化が起こる兆しを発見し、実態に変化が起る前に対策をとる。	
② 異常検出型	不正検出型	大量のデータの中から異常なパターンを自動的にクエリ/タイムに検出し、早期の危機対応を促す。
	外れ値検出型	手前の知識による異常パターンの位置が容易な場合には、異質(不正)の定義を作成し、これに合致する行動や状態などを不正と見なして検出する。
③ 予測・制御型	収益シミュレーション型	業務における収益発生メカニズムをモデル化し、想定する業務改善施策の潜在効果をシミュレーションにより試算し、施策の優先づけや、基盤を行う。
	リスク・シミュレーション型	不測な要素に起因するリスクを考慮したシミュレーションを行い、業務改善施策の優先づけや、基盤を行う。
	最適化型	最適化手法を用いて、業務のパフォーマンスを最大化するよう改善施策を提案する。
	リスク・ヘッジ型	リスクを定量化し、そのリスクを最小にするような業務改善施策を提案する。
④ ターゲティング型	顧客層や見込み顧客など、熟練者から考えて単点的に抽出するべきターゲットを抽出することにより、アクションの費用対効果を最大化する。	
⑤ 与信管理型	顧客のデフォルト(滞債・倒産)リスクや解約リスクをスコアリングすることによって、顧客の選別や課税優先順位を打つ。	
⑥ 評価・要因分析型	課題対象を比較評価すると同時に、評価結果を左右する要因を特定し、対策ごとに評価の改善施策を提示する。	
⑦ コンテキスト・アウェアネス型	個々のユーザーの行動履歴や嗜好などのデータから、行動の意図を読み取り、一歩先回りした商品やサービスを提供する。	
⑧ プロセス・トレース型	顧客の優良顧客化や苦痛の軽減化など、成長・継続や他社のプロセスを抽出し、それを促進または阻害・阻止する要因を特定する。	

表2 分析シナリオ類型の典型的な事例

分析シナリオ類型	代表的な事例	
① 予兆発見型	・ インターネットの利用傾向から、商品やサービスの普及の兆しを発見する。	
② 異常検出型	不正検出型	・ 申請書類 集約において不正や誤りを含む申請を検出し、注意を喚起する。
	外れ値検出型	・ 不正購入傾向システムにおいて特異なアクセスを不正検出として検出し、注意を喚起する。
③ 予測・制御型	収益シミュレーション型	・ 健康増進施策を実施した場合の医療費削減効果を試算する。
	リスク・シミュレーション型	・ ホテルの客室や生鮮食品の値引きのタイミング調整による売れ残りリスクを試算する。
	最適化型	・ コルセナにおける頻率的なオペレータの配置計画や勤務スケジュールを決定する。
④ ターゲティング型	リスク・ヘッジ型	・ 洪水備わつて定めた効果が得られる広告費の配分を決定する。
	④ ターゲティング型	・ 顧客の属性や購買履歴から優良顧客や見込み顧客を抽出し、ダイレクトメールの送信対象とする。 ・ レビノおよび健康診断データから生活習慣病の予備群を抽出し、健康増進施策の実施対象とする。
⑤ 与信管理型	・ 過去の債務不履行データを基に、個人の滞債リスクや企業の倒産リスクをスコアリングする。 ・ 貸付サービスにおける顧客の属性や利用履歴から、解約リスクをスコアリングする。	
⑥ 評価・要因分析型	・ 営業店舗や営業員の成績を多面的に評価し、成績を向上させる要因を特定する。 ・ 費用対効果の観点から顧客を客群的に評価し、価格調整に当たっての方針を策定する。	
⑦ コンテキスト・アウェアネス型	・ ユーザーの利用履歴や現在の利用状況から嗜好を読み取り、商品のレコメンデーションやサービスのパーソライゼーションを実施する。	
⑧ プロセス・トレース型	・ ECサイトにおける来店から購買に至るプロセスを分析し、促進要因またはボトルネック要因を特定する。 ・ 疾病の重症化プロセスを抽出し、これを阻止する要因を特定する。	

なお、(株)数理システムの Visual Mining Studio (以下、VMS) は、データマイニング・ツールとして広く知られているが、我々はそれに加えて、分析履歴の保存や担当者間における分析手順の共有を可能にする知識共有ツールとしての VMS に着目しており、さきに紹介した「分析シナリオ類型」もこの VMS を用いてテンプレート化している。

テンプレートとは、分析シナリオごとに典型的な事例の標準的な分析フローを VMS により実装したものであり(図3参照)、ここで VMS はシナリオのエディタとしての役割を果たしている。分析フローを構成する手法には、データマイニングや統計分析だけでなく、シミュレーションや最適化など様々なものが含まれるため、VMS から各種の計算エンジン( DMExpress[5] , Visual SLAM[6] , NUOPT[7]など) を呼び出すためのアドオンを必要に応じて開発し、一連の分析作業をシームレスに実行できるようにしている。

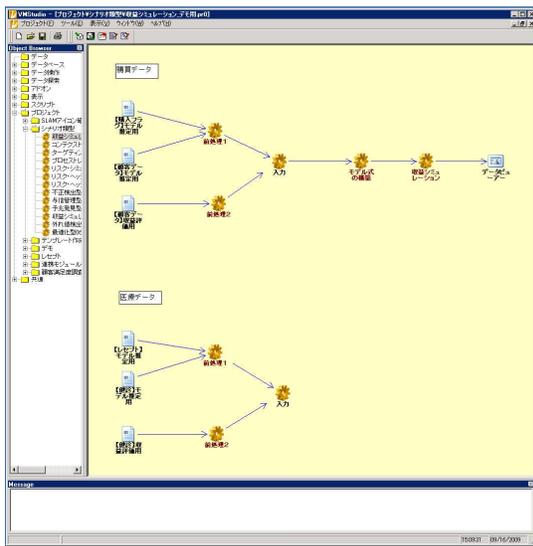


図3 分析テンプレートのイメージ  
(収益シミュレーション型)

#### 4 おわりに

今回紹介したレセプト・健診データ分析は、疾病予防と医療費削減を目指す保健事業の実施に際して、保健指導人材の配分先を重点化することによって費用対効果を高めようとするもの

であった。そして、これと非常に似通った分析を過去にマーケティング目的でも実施していたことが、分析シナリオ類型の整備に思い至った動機である。VMS を用いて分析シナリオ類型のテンプレート化を進めるうちに、データマイニングや統計分析だけでなくシミュレーションや最適化のエンジンも VMS から自由に呼び出し、手法を越えたシームレスな分析を実現したいという要望が生じ、その後の各種アドオンの開発にもつながっている。VMS の更なる発展に期待したい。

#### 参考文献

- [1] 厚生労働省大臣官房統計情報部：“平成18年度国民医療費の概況”，  
<http://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/06/index.html>
- [2] 別府洋美：“レセプト（レセプトのオンライン請求義務化）とは”，  
<http://itpro.nikkeibp.co.jp/article/Keyword/20071010/284186/>
- [3] M. J. Zaki：“SPADE: an efficient algorithm for mining frequent sequences,” Machine Learning Journal Vol.42, 2001.
- [4] M. Ben-Akiva and S. R. Lerman：“Discrete Choice Analysis: Theory and Application to Travel Demand,” The MIT Press, 1985.
- [5] 株式会社アシスト：“高速データ加工ツール Syncsort DMExpress”，  
<http://www.ashisuto.co.jp/prod/syncsort/>
- [6] 株式会社構造計画研究所：“汎用シミュレーションモデル構築ツール Visual SLAM”，  
<http://www4.kke.co.jp/orsim/visualslam/>
- [7] 株式会社数理システム：“数理計画法パッケージ NUOPT”，  
<http://www.msi.co.jp/nuopt/>
- [8] 高橋彰子, 山下友子, 濱宏一郎, 中川慶一郎：“根拠に基づく保健事業(EBH)を支援する疾病管理プログラム”, オペレーションズ・リサーチ Vol.54, 2009.
- [9] 株式会社 NTT データ技術開発本部ビジネスインテリジェンス推進センター：“BI 革命”, NTT 出版, 2009.