

一般口演

一般口演7

医療データ分析3（レセプトデータ・治験）

2018年11月23日(金) 14:20 ～ 15:50 H会場 (福岡サンパレスHパレスホール)

[2-H-2-1] 電子レセプト情報の分析のための解析論理の記述方式に関する検討

○佐藤 淳平¹, 合田 和生¹, 喜連川 優¹, 満武 巨裕² (1.東京大学 生産技術研究所, 2.医療経済研究機構)

電子レセプト情報は、元来、診療報酬請求の為の情報であり、電子カルテ等と比較すると医療に関する限定的な情報に過ぎないと言えるものの、我が国の医療制度が公的医療保険を基礎としており、国内で提供されているほぼ全ての医療に関して電子レセプト情報が生成されていることを鑑みると、当該情報を分析することは、今後の医療政策の立案に大いに資するものを期待されている。著者らは、過去、大規模な電子レセプト情報の提供を受け、当該情報を解析する為の高速データベース基盤を構築すると共に、当該基盤を元とした解析サービスを研究者らに提供してきた。この過程で、医療分野の研究者と共に、電子レセプト情報の分析の為の解析論理を検討する機会を多々得たが、電子レセプト情報は医療分野の研究者に馴染みのない順序付き入れ子タプル構造を有し、当該構造を前提として解析論理を簡便かつ明瞭に記述する方式が未だ確立されていないことから、医療分野の研究者が分析に関する考案を論理的に記述することに困難が生じ、また、多くの時間を要していることが判った。本研究では、電子レセプト情報の分析をより活性化するべく、簡便かつ明瞭な解析論理の記述方式を確立することを目指し、電子レセプト情報が備えるデータ構造ならびに当該情報に対する典型的な演算について、形式的な記述方式を整理して提案する。また、分析の事例を提示し、当該方式による記述例を示す。加えて、解析論理の処理系の在り方を論ずる。

電子レセプト情報の分析のための解析論理の記述方式に関する検討

佐藤 淳平*1、合田 和生*1、喜連川 優*1、満武 巨裕*2

*1 東京大学 生産技術研究所、*2 医療経済研究機構

A Preliminary Study on Description Scheme of Analytics Logic for Health Insurance Claims Data

Jumpei Sato*1, Kazuo Goda*1, Masaru Kitsuregawa*1, Naohiro Mitsutake*2

*1 Institute of Industrial Science, The University of Tokyo, *2 Institute for Health Economics and Policy

Health insurance claims are invoices of medical service fees for insured medical services provided to patients. The claims data is generated for almost all medical services under the universal health insurance system in Japan. Japan's Ministry of Health, Labour and Welfare (MHLW) has been continuously collecting the claims data in order to develop, implement and evaluate medical care expenditure regulation plans. We have been provided with a 6-year data set of the claims data from MHLW and have offered analytics services of the claims data to researchers. Many researchers have needed much time to define and describe their analytics ideas because a scheme for easily and clearly describing an analytics logic for the claims data has not been established. In this work, we propose a description scheme of analytics logic for the claims data, and demonstrate the effectiveness of the proposed description scheme by showing the execution results of typical analytics.

Keywords: Health Insurance Claims Data, Description Scheme, Analytics Logic.

1. 緒論

厚生労働省は、「高齢者の医療の確保に関する法律」が制定された 2008 年 4 月の翌年から今日に至るまで、全国医療費適正化計画及び都道府県医療費適正化計画の作成、実施及び評価を目的として、国内全ての保険者等から匿名化された電子レセプト情報を継続的に収集し、収集した電子レセプト情報に基づき「レセプト情報・特定健診等情報データベース(通称 NDB)」の構築を行っている¹⁾。電子レセプト情報は医療機関が患者に提供した保険診療に対する医療報酬を保険者に請求する為の情報であり、公的医療保険制度の下、我が国に於けるほぼ全ての医療行為に対して電子レセプト情報が生成されている。当該データベースを分析することは、医療費適正化計画の作成、実施及び評価に加え、他の医療政策の立案に大いに貢献するものと期待されている。

著者らは、厚生労働省より当該データベースから 6 年分の電子レセプト情報の提供を受け、当該情報の分析を可能とする高速データベース基盤を構築すると共に、当該基盤を元とした解析サービスを国内の医療や政策立案分野の実務者や研究者らに提供してきた²⁾³⁾⁴⁾。この過程で、当該実務者や研究者らと共に、電子レセプト情報の分析の為の解析論理を検討する機会を多々得たが、当該情報が有するデータ構造を前提とし解析理論を簡便かつ明瞭に記載する方法が未だ確立されていないため、当該実務者や研究者が分析に関する新たな考案を論理的に記述する事に困難を生じ、また、多くの時間を要していることを痛感した。

2. 目的

本論文では、全国医療費適正化計画及び都道府県医療費適正化計画の作成、実施及び評価をより機動的に実行可能とするべく、電子レセプト情報の簡便かつ明瞭な解析論理の記述方式を確立することを目指し、予備的な検討として当該情報を用いた分析に於いて典型的に実行される解析処理の記述方式を整理して提案する。また、電子レセプト情報を用いた典型的な分析例を提示し、当該方式による解析論理の記述例と、解析論理に基づく分析の結果を示す。加えて、解析論理の処理系の在り方を論ずる。

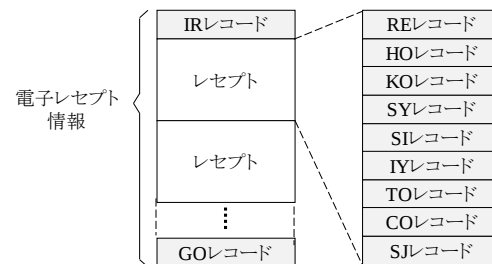


図1 電子レセプト情報のデータ構造の例

3. 方法

3.1 解析論理の記述方式

電子レセプト情報は、従前の紙媒体を用いた診療報酬の請求手続きに於いて使用されていた記載様式を踏襲したデータ構造を有している。電子レセプト情報を格納するファイルはいわゆる CSV (comma-separated values) 形式のテキストにより構成され、テキスト中の各行はレコードと呼ばれる⁵⁾。各レコードの 1 番目のカラム(英数字 2 文字)に記載される情報が、そのレコードの種別を表し、レコード種別ごとに各カラムに記載される情報は異なり、RE レコードを先頭とする一連のレコード群が 1 件のレセプト(claim)を構成する(図 1)。即ち、電子レセプト情報は順序付き入れ子ダブル構造によって構成される。代表的なレコード種別である SY レコードには傷病に関する情報が記載され、当該レコードの 2 番目のカラムには傷病名コード、3 番目のカラムには傷病の診療開始日がそれぞれ記載される。

他のビッグデータ解析と同様に、電子レセプト情報の解析に於いても、殆どの分析課題は過去試みられていないアドホックなものである。即ち、既存の解析ツールで直接求解可能な分析課題は限られており、分析課題に基づき解析論理を設計し、これを実装するプロセスが欠かせない。当該プロセスは、医療・政策立案分野の研究者・実務者と情報学分野の研究者が共同で行うこととなるが、電子レセプト情報の解析論理をフォーマルに記述する方式が確立されていないことから、

分析に係る考案を医療・政策立案分野の研究者・実務者が明瞭に記述することに困難が生じているのが実情である。故に、分析チーム内に於ける解析論理の議論や、分析に使用する解析論理の資産化を行う事ができず、プロセスが場当たり的となる傾向が強く、非効率的に時間を要している。本論文では、分析チーム内での解析論理の議論や、分析に使用する解析論理の資産化を可能とし、分析プロセスを効率化すべく、電子レセプト情報に対する解析論理を記述する方式を提案する。

まず、電子レセプト情報の解析に於けるデータ型を以下のように定義する。

- ・ c: 複数のレコードから構成される単一のレセプト
- ・ e: 複数のレセプトから構成される単一のエピソード
- ・ t: 複数の値から構成される単一のタプル
- ・ V: 単一の値

患者の経時的な病態の変化を分析する分析事例では、複数期間に発行されたレセプトを患者 ID 別に分類し、分類された複数のレセプトに基づき患者の病態を分類する解析処理が必要であることが多い。また、解析処理の結果は最終的に既存の分析システムで集計可能なタプル形式のデータに変換され、当該タプルデータを用いた各種統計値の算出処理が実施される傾向が強い。レセプトを表す c に加えて、特定の値に基づき分類した複数のレセプトを表すデータのデータ型を e (エピソード) と、複数の値から構成されるリレーショナルタプル形式のデータを t (タプル) と、単一の値 V (バリュー) を用いて、これらの処理を記述する事を可能とする。

また、各データ型で定義されるデータの集合型として、レセプト c の集合型 C、エピソード e の集合型 E、タプル t の集合型 R をそれぞれ定義する。集合型を定義することにより、データ型が同一である複数のデータに対する処理を記述することを可能とする。なお、各データ型で定義されるデータのオブジェクトは $c_n, e_n, t_n, V_n, C_n, E_n, R_n$ と記述する (n は整数)。

更に、電子レセプト情報を用いた分析に於いて典型的に使用される解析処理をデータ処理関数として定義する (図 2)。定義するデータ処理関数の概要は以下の通りである。

- ・ $f_{rez_to_tup}$: 単一のレセプトの集合型 C を入力とし、単一のタプルの集合型 R を出力する
- ・ $f_{rez_to_ep}$: 単一のレセプトの集合型 C を入力とし、単一のエピソードの集合型 E を出力する
- ・ $f_{resz_to_rez}$: 2 つの異なるレセプトの集合型 C を入力とし、単一のレセプトの集合型 C を出力する
- ・ $f_{ep_to_tup}$: 単一のエピソードの集合型 E を入力とし、単一のタプルの集合型 R を出力する
- ・ $f_{eps_to_ep}$: 2 つの異なるエピソードの集合型 E を入力とし、単一のエピソードの集合型 E を出力する
- ・ $f_{tup_to_tup}$: 単一のタプルの集合型 R を入力とし、単一のタプルの集合型 R を出力する

各関数では、入力データから出力データの生成に必要な典型的な演算について処理内容を定義する。各演算の処理内容を以下に示す。なお、各処理内容の記載に於いて、添え字 n, m, i, j は値の異なる整数を意味する。

- ・ **Selection:** レセプトの集合 C_n またはエピソードの集合 E_n を入力とし、 C_n または E_n に含まれる c_n または e_n のうち特定の条件を満たす c_n または e_n のみから構成される C_m または E_m を出力する。例えば、75 歳以上の患者のレセプトから構成されるレセプトの集合を出力する場合、 $c.RE_{<age>} = 75$ を条件として定義する

$$\begin{aligned} f_{rez_to_tup} &: C \xrightarrow{\text{selection}} C \xrightarrow{\text{conversion}} C \xrightarrow{\text{extraction}} R \\ f_{rez_to_ep} &: C \xrightarrow{\text{selection}} C \xrightarrow{\text{conversion}} C \xrightarrow{\text{sessionization}} E \\ f_{resz_to_rez} &: C \times C \xrightarrow{\text{reduction}} C \\ f_{ep_to_tup} &: E \xrightarrow{\text{selection}} E \xrightarrow{\text{conversion}} E \xrightarrow{\text{summarization}} R \\ f_{eps_to_ep} &: E \times E \xrightarrow{\text{reduction}} E \\ f_{tup_to_tup} &: R \xrightarrow{\text{selection}} R \xrightarrow{\text{sql-based-conversion}} R \end{aligned}$$

図 2 提案方式に於いて使用するデータ処理関数の定義

- ・ **Conversion:** レセプトの集合 C_n またはエピソードの集合 E_n を入力とし、 C_n に含まれる c_n または E_n に含まれる e_n を構成する c_n に対して特定の算術処理を行い、算術処理後の c_n より構成される C_m または E_m を出力する。例えば、各レセプトに記載される患者の年齢の値に 1 を加える処理を行う場合、 $c.RE_{<age>} = c.RE_{<age>} + 1$ を算術処理として定義する
- ・ **Extraction:** レセプトの集合 C_n を入力とし、 C_n に含まれる各 c_n を t_n に変換し、 t_n より構成されるタプルの集合 R_n を出力する。なお、 t_n を構成する値 V_n は c_n に含まれる情報をもとに算出される。例えば、各レセプトに記載される患者 ID の値を V_1 として抽出する場合、 $t.V_1 = c.RE_{<patient_id>}$ を定義する
- ・ **Sessionization:** レセプトの集合 C_n を入力とし、 C_n に含まれる c_n を特定の集約 key を用いて集約した e_n を作成し、 e_n より構成されるエピソードの集合 E_n を出力する。例えば、患者 ID を集約 key とする場合、 $key_1 = c.RE_{<patient_id>}$ を定義する
- ・ **Merge (×):** 2 つの異なるレセプトの集合 C_n, C_m または 2 つ異なるエピソードの集合 E_n, E_m を入力とし、 C_n に含まれる c_n と C_m に含まれる c_m を特定の条件に従い結合した c_i 、または E_n に含まれる e_n と E_m に含まれる e_m を特定の条件に従い結合した e_i を作成し、 c_i より構成される C_i または e_i より構成される E_i を出力する。例えば、 E_1 と E_2 を入力とし、 E_1 と E_2 に含まれるエピソードのうち、同一の患者 ID が記載されているレセプトを有するエピソードを結合する場合、 $E_1.e.c[0].RE_{<patient_id>} = E_2.e.c[0].RE_{<patient_id>}$ を定義する
- ・ **Reduction:** 2 つの異なるレセプトの集合 C_n, C_m の結合結果 C_i または 2 つの異なるエピソードの集合 E_n, E_m の結合結果 E_i を入力とし、 C_i に含まれる c_i または E_i に含まれる e_i のうち特定の条件を満たす c_i または e_i のみから構成される C_j または E_j を出力する。例えば、 E_1 と E_2 の結合結果に含まれるエピソードのうち、 E_1 由来のエピソードに含まれるレセプトのいずれかに糖尿病の傷病名コードの集合 {DM_DISEASE_CODES} 内のいずれかの傷病名コードが記載されているエピソードより構成されるエピソードの集合を出力する場合、 $(\exists c \in e) (\exists r \in c.SY[]) (r.disease_code \in \{DM_DISEASE_CODES\})$ を定義する
- ・ **Summarization:** エピソードの集合 E_n を入力とし、 E_n に含まれる各 e_n を t_n に変換し、 t_n より構成されるタプルの集合 R_n を出力する。なお、 t_n を構成する値 V_n は e_n を構成する c_n に含まれる情報をもとに算出される。例えば、各エピソードに含まれるレセプトのいずれかに記述されている患者 ID の値を V_1 として抽出する場合、 $t.V_1 = e.c[0].RE_{<patient_id>}$ を定義する
- ・ **SQL-based-conversion:** タプルの集合 R_n を入力とし、 R_n に含まれる t_n に対して特定の集計処理を行い、集計処理

の結果として得られるタプル t_m より構成される R_m を出力する。例えば、値 V_1 , V_2 より構成されるタプルの集合を入力とし、 V_2 を集計軸とし V_1 の合計値を算出する場合、 $\text{SELECT SUM}(V_1), V_2 \text{ GROUP BY } V_2$ を定義する

過去の電子レセプト情報を用いた分析は、多くの場合、特定の条件により抽出したレセプトのエピソードへの変換処理 ($f_{rez_to_ep}$)、特定の条件により抽出したレセプトまたはエピソードのタプルへの変換処理 ($f_{rez_to_tup}$, $f_{ep_to_tup}$)、2つの異なる条件により抽出したレセプトまたはエピソードの結合処理 ($f_{res_to_res}$, $f_{eps_to_ep}$)、タプルを用いた各種統計値の算出処理 ($f_{tup_to_tup}$) の組合せによって実行可能である。そのため、全ての解析欲求に応えることは不可能であるが、電子レセプト情報を用いた典型的な分析に対して、上記処理関数を組み合わせた処理の手順と、各関数に於ける処理内容を定義することにより、解析論理の記述を可能とする。

一般的に電子レセプト情報はレセプト発行元の医療機関種別に応じて記載様式が定められており、医療機関種別は主に、クリニックまたは一般病院 (MED)、包括支払い制度対象病院 (DPC)、歯科医院 (DEN)、調剤薬局 (PHA) の4種類に分類される⁷⁾。レセプト発行元の医療機関種別を表す情報はレセプト内に記載されていないが、例えば、特定の医療機関より発行されたレセプトのみを対象とする分析では、処理中に各レセプトの発行元の医療機関種別の情報を参照する必要がある。本論文では、分析処理に於いて各レセプトの発行元の医療機関種別を参照可能とするため、医療機関種別を表す情報を RE レコードの <rezept type> カラムに記載する。

3.2 解析論理に基づく分析

提案方式の有効性を示すため、著者らがこれまでに医療や政策立案分野の実務者や研究者と共に行った分析のうち、典型的な3種類の分析例について、提案方式による解析論理の記述例と、解析論理に基づく分析の実行結果を示す⁵⁾⁶⁾。

まず、各レセプトに記載された情報に基づき集計処理を行う分析の例として、「特定年度に於ける年齢階層別の一人当たり医療費の分析」の結果を示す。次に、特定の値が等しいレセプトより作成したエピソードに基づき患者の病態の解析を行う分析の例として、「年度別の糖尿病・高血圧・高脂血症患者数の分析」の結果を示す⁵⁾⁶⁾。最後に、特定の値が等しいレセプトより作成したエピソードに基づき、患者の病態の経時的な変化の解析を行う分析の例として、「非糖尿病患者群の重症化推移の分析」の結果を示す⁵⁾⁶⁾。

それぞれの分析例について、解析論理の記述内容に則り処理を実行するプログラムを作成し、著者らがこれまでに構築した高速データベース基盤上にて当該プログラムを実行することで分析を行う。分析例のうち、「特定年度に於ける年齢階層別の一人当たり医療費の分析」と「年度別の糖尿病・高血圧・高脂血症患者数の分析」は、厚生労働省より提供を受けた6年分の匿名化済みの電子レセプト情報を使用して分析を行う。「非糖尿病患者群の重症化推移の分析」については、これまでの解析サービスを提供する取り組みの中で三重県地域の保険者より提供を受けた匿名化済みの電子レセプト情報を使用して分析を行う。

4. 結果

4.1 特定年度に於ける年齢階層別の一人当たり医療費の分析

当該分析の解析論理の記述例を図3に示す。図3の1行

目に示すように、当該分析では $f_{rez_to_tup}$, $f_{tup_to_tup}$ を使用し分析を行った。解析処理の流れは以下の通りである。(1) $f_{rez_to_tup}$ を使用し、診療年月の年度が2014年のレセプトより、患者ID、レセプト発行元の医療機関種別、患者の年齢階層、医療費の値を抽出し、抽出した値より構成されるタプルを作成した。医療機関種別がMEDである場合、レセプト種別情報に基づき、医療機関種別を医科外来レセプト (MED-out-patient) または医科入院レセプト (MED-in-patient) に分類した。(2) 次に、 $f_{tup_to_tup}$ を使用し、各タプルより年齢階層と医療機関種別を集計軸として医療費の合計値とユニークな患者ID数を算出し、医療費の合計値をユニークな患者ID数で除算することにより一人当たり医療費を算出した。

当該処理の実行結果を元に作成したグラフを図4示す。本結果から、当該方式を用いることで、各レセプトに記載された情報に基づき集計処理を行う分析の解析論理の記述が可能であり、当該記述に則り分析を実行させることが可能なことを確認した。

4.2 年度別の糖尿病・高血圧・高脂血症患者数の分析

当該分析に於ける解析論理の記述例を図5に示す。図5に於いて、{DM_DISEASE_CODES} と {DM_DRUG_CODES} は糖尿病の傷病名コードと治療薬剤の薬剤コードの集合、{HBP_DISEASE_CODES} と {HBP_DRUG_CODES} は高血圧の傷病名コードと治療薬剤の薬剤コードの集合、{HL_DISEASE_CODES} と {HL_DRUG_CODES} は高脂血症の傷病名コードと治療薬剤の薬剤コードの集合をそれぞれ意味する。図5の1行目に示すように、当該分析では $f_{rez_to_ep}$, $f_{ep_to_tup}$, $f_{tup_to_tup}$ を使用し分析を行った。各関数を使用した解析処理の流れは以下の通りである。(1) $f_{rez_to_ep}$ を使用し、医療機関種別がPHAであるレセプト、または医療機関種別がMEDかつレセプト種別が外来レセプト、即ち医科外来レセプト (MED-out-patient) のレセプトを対象とし、診療年月の年度と患者IDを集約keyとしてエピソードを作成した。(2) 次に、 $f_{ep_to_tup}$ を使用することで各エピソードに含まれるレセプトより、患者IDと診療年月の会計年度、糖尿病、高血圧、高脂血症の罹患状態を表す値を抽出し、抽出した値より構成されるタプルを作成した。(3) そして $f_{tup_to_tup}$ を用いて、各タプルより診療年月の会計年度と糖尿病、高血圧、高脂血症の罹患状態を集計軸として患者数の総和を算出した。

当該処理の実行結果を元に作成したグラフを図6示す。本結果から、当該方式を用いることで、特定の値が等しいレセプトより作成したエピソードに基づき患者の病態を分類する解析を行う分析の解析論理の記述が可能であり、当該記述に則り分析を実行させることが可能なことを確認した。

4.3 非糖尿病患者群の重症化推移の分析

当該分析に於ける解析論理の記述例を図7に示す。図7に於いて、{DM_DISEASE_CODES} は糖尿病の傷病名コード、{DM_DRUG_CODES} は糖尿病の治療薬剤の薬剤コード、{DIALYSIS_CODES} は透析の診療行為コードのそれぞれ集合を意味する。図7の1行目に示すように、当該分析では2つの $f_{rez_to_ep}$ ($f_{rez_to_ep1}$ および $f_{rez_to_ep2}$)、 $f_{eps_to_eps}$, $f_{ep_to_tup}$, $f_{tup_to_tup}$ を使用し分析を行った。各関数を使用した解析処理の流れは以下の通りである。(1) $f_{rez_to_ep1}$ を使用し、診療年月が2013/04から2013/09の間であり且つ年齢が75歳以上の患者のレセプトを対象とし、診療年月の年度と患者IDを集約keyとしてエピソードを作成した (E_1)。 (2) $f_{rez_to_ep2}$ を使用し、診

療年月が 2013/10 から 2016/09 の間であるレセプトを対象とし、診療年月から算出した期間情報 (2013/10-, 2014/04-, 2014/10-, 2015/04-, 2015/10-, 2016/04-) と患者 ID を集約 key としてエピソードを作成した (E_2)。 (3) 次に、 $f_{eps_to_ep}$ を使用し、 E_1 に含まれるのエピソードのうち糖尿病の疾患コードと治療薬のコードを含むレセプトを有さないエピソード (即ち非糖尿病患者のエピソード) と E_2 に含まれるエピソードを、患者 ID を key として結合した。 (4) $f_{ep_to_tup}$ を使用し、結合した各エピソードより、患者 ID、診療年月より算出した期間情報 (2013/10-, 2014/04-, 2014/10-, 2015/04-, 2015/10-, 2016/04-)、糖尿病の罹患ステージ情報を表す値を抽出し、抽出した値より構成されるタプルを作成した。 (5) そして $f_{tup_to_tup}$ を用いて、各タプルより糖尿病の罹患ステージと算出した期間情報を集計軸とし患者数の総和を算出した。

当該処理の実行結果を元に作成したグラフを図 8 示す。本結果から、当該方式を用いることで、特定の値が等しいレセプトより作成したエピソードに基づき、患者の病態の経時的な変化の解析を行う分析の解析論理の記述が可能であり、当該記述に則り分析を実行させることが可能なことを確認した。

このように、当該方式を使用することにより、電子レセプト情報を使用した典型的な 3 種類の分析について、簡便かつ明瞭に解析論理を記述可能であり、当該解析論理に則り分析を実行させることが可能であることを確認できた。

5. 考察

本論文では、電子レセプト情報の簡便かつ明瞭な解析論理の記述方式を確立することを目指し、予備的な検討として当該情報を用いた分析に於いて典型的に実行される解析処理の記述方式を整理して提案した。更に、当該情報を用いた典型的な 3 種類の分析例について解析論理の記述例を示し、解析論理に基づき分析を実行させることが可能なことを示した。1 つ目の分析例「特定年度に於ける年齢階層別の一人当たり医療費の分析」で用いた処理により、国家規模や特定地域に於ける医療費や医療資源の投入量等の統計値を算出する分析が実行可能である。2 つ目の分析例「年度別の糖尿病・高血圧・高脂血症患者数の分析」で用いた処理により、例えば、任意の疾患の罹患状態の組合せについて患者数を算出する分析が実行可能である。3 つ目の分析例「非糖尿病患者群の重症化推移の分析」で用いた処理により、任意の疾患について、患者の病態の経時的な変化に関する分析が実行可能である。以上より、提案方式を使用することで電子レセプト情報を用いた多くの分析が簡便かつ明瞭に記述することが可能であり、解析論理に則り分析が実行可能であることを確認した。

今後の課題は、提案方式の有効性の検証と、提案方式により記述した解析処理に則り、容易に分析が実行可能な分析システムの開発である。本論文では典型的な分析例を用いて提案方式の有効性を確認したが、当該方式の有効性の検証には、分析の完遂に要する時間の短縮効果等の検証が必要である。また、提案方式により記述した解析論理を簡便かつ明瞭に指定可能な User interface を有する分析システムの開発することで、解析プログラムの構築を行わずに分析の実行が可能となり、より機動的に分析を行う事が可能となる。一般的な分析システムでは、分析に使用するデータはリレーショナルデータベースにレコード単位の形式で格納され、レコード単位で処理が実行されることが多い。しかしながら、当該記述方式では、主にレセプト単位で解析処理を記述する。その為、提案する記述方式に基づき処理を実行する分析システ

ムでは、レセプト単位でデータを格納することで、効率的に処理が実行可能であると考えられる。今後、前述の分析システムの開発を行い、多くの研究者らに使用頂くことで、全国医療費適正化計画及び都道府県医療費適正化計画の作成、実施及び評価をより機動的に実行可能としてゆく。

6. 結論

本論文では、電子レセプト情報の簡便かつ明瞭な解析論理の記述方式を確立することを目指し、予備的な検討として当該情報を用いた分析に於いて典型的に実行される解析処理の記述方式を整理して提案した。当該記述方式を用いることで、当該情報を使用した典型的な 3 種類の分析について、簡便かつ明瞭に解析論理の記述方式を記載することが可能であり、また、解析論理に則り分析が実行可能であることを示し、当該記述方式の有効性を確認した。

謝辞

本研究の一部は、厚生労働科学研究費政策科学推進研究「汎用性の高いレセプト基本データセット作成に関する研究」、厚生労働科学特別研究事業戦略研究「レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビッグデータ解析基盤の整備」、内閣府最先端研究開発支援プログラム (FIRST) 「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」、内閣府革新的研究開発推進プログラム (ImPACT) 「社会リスクを低減する超ビッグデータプラットフォーム」、日本医療研究開発機構 (AMED) 臨床研究等 ICT 基盤構築研究事業「エビデンスの飛躍的創出を可能とする超高速・超学際次世代 NDB データ研究基盤構築に関する研究」の助成に依る。レセプト情報・特定健診等情報データベースからの電子レセプト情報の第三者提供に掛かる手続きに関しては、厚生労働省保険局保険システム高度化推進室から丁寧なご指導を頂いた。

参考文献

- 1) 厚生労働省. 健康保険法等の一部を改正する法律について.
[<https://www.mhlw.go.jp/bunya/iryouhoken/iryouhoken02/01.html> (2018-08-17)]
- 2) 厚生労働省老健局・保険局.NDB、介護 DB 等の役割と解析基盤について. 2018.
[<https://www.mhlw.go.jp/content/12401000/000332845.pdf> (2018-08-17)]
- 3) 満武巨裕. 日本のレセプト情報・特定健診等データベース (NDB) の有効活用. 情報処理, 2015, Vol. 56, No. 2, pp. 140-144.
- 4) 合田和生, 山田浩之, 喜連川優, 満武巨裕. 我が国の公的医療保険の悉皆分析を可能とする高速レセプト解析システムの開発と今後の展望. 第 15 回日本データベース学会年次大会 (DEIM2017), 2017.
- 5) 中島直樹, 朴珍相. NDB を用いて、日本の生活習慣病の全体像をいかに正確に把握するか. 日本医療情報学会合同シンポジウム: 世界最大級の医療 Real World Data, NDB を用いた糖尿病研究. 第 61 回日本糖尿病学会年次学術集会 (平成 30 年 5 月 24 日)
- 6) 満武巨裕. エビデンスの飛躍的創出を可能とする超高速・超学際次世代 NDB データ研究基盤構築を活用した糖尿病関連エビデンスと課題. 日本医療情報学会合同シンポジウム: 世界最大級の医療 Real World Data, NDB を用いた糖尿病研究. 第 61 回日本糖尿病学会年次学術集会 (平成 30 年 5 月 24 日)
- 7) 厚生労働省保険局. 診療報酬情報提供サービス
[<http://www.iryohoken.go.jp/shinryohoshu/> (2018-08-17)]

Function flow definition: $input \xrightarrow{f_{rez.to.tup}} R_1 \xrightarrow{f_{tup.to.tup}} output$

Parameters of $f_{rez.to.tup}$:

selection: $c \mid (c \in C)(FiscalYear(c.RE_{<intervention_year_and_month>}) = 2014)$ conversion: none

extraction:

$t.V_1 = c.RE_{patient.id}$ $t.V_2 = c.RE_{age.category}$

$t.V_3 = \begin{cases} \text{'MED out-patient'} & \text{if } c.RE_{<rezept_type>} = \text{'MED'} \text{ AND } c.RE_{<rezept_kind>} \%2 = 0 \\ \text{'MED in-patient'} & \text{if } c.RE_{<rezept_type>} = \text{'MED'} \text{ AND } c.RE_{<rezept_kind>} \%2 = 1 \\ c.RE_{<rezept_type>} & \text{otherwise} \end{cases}$

$t.V_4 = \sum_{r \in c.HO[]}$ $\begin{cases} r_{<total_points>} & \text{if } c.RE_{<rezept_type>} \in \{\text{'MED'}, \text{'DEN'}, \text{'PHA'}\} \text{ OR} \\ & (c.RE_{<rezept_type>} = \text{'DPC'} \text{ AND } r_{<total_point_record_flag>} = 1) \\ 0 & \text{otherwise} \end{cases}$

Parameters of $f_{tup.to.tup}$:

selection: none sql-based-conversion: **SELECT** $V_2, V_3, \text{SUM}(V_4)/\text{COUNT}(\text{DISTINCT } V_1)$ **GROUP BY** V_2, V_3

図3 特定年度に於ける年齢階層別の一人当たり医療費の分析に於ける解析論理の記述例

Function flow definition: $input \xrightarrow{f_{rez.to.ep}} E_1 \xrightarrow{f_{ep.to.tup}} R_1 \xrightarrow{f_{tup.to.tup}} output$

Parameters of $f_{rez.to.ep}$:

selection:

$c \mid (c \in C)(c.RE_{<receipt_kind>} = \text{'MED'} \text{ AND } c.RE_{<receipt_kind>} \%2 = 0) \text{ OR } (c \in C)(c.RE_{<receipt_kind>} = \text{'PHA'})$

conversion: none

sessionize:

$key_1 = FiscalYear(c.RE_{<intervention_year_and_month>})$ $key_2 = c.RE_{<patient.id>}$

Parameters of $f_{ep.to.tup}$:

selection: none conversion: none

summarization:

$t.V_1 = e.c[0].RE_{<patient.id>}$

$t.V_2 = FiscalYear(e.c[0].RE_{<intervention_year_and_month>})$

$t.V_3 = \begin{cases} \text{'DM(+)} & \text{if } (\exists c \in e)(\exists r \in c.SY[])(r_{<disease_code>} \in \{\text{DM_DISEASE_CODES}\}) \text{ AND} \\ & (\exists c \in e)(\exists r \in c.IY[])(r_{<drug_code>} \in \{\text{DM_DRUG_CODES}\}) \\ \text{'DM(-)} & \text{otherwise} \end{cases}$

$t.V_4 = \begin{cases} \text{'HBP(+)} & \text{if } (\exists c \in e)(\exists r \in c.SY[])(r_{<disease_code>} \in \{\text{HBP_DISEASE_CODES}\}) \text{ AND} \\ & (\exists c \in e)(\exists r \in c.IY[])(r_{<drug_code>} \in \{\text{HBP_DRUG_CODES}\}) \\ \text{'HBP(-)} & \text{otherwise} \end{cases}$

$t.V_5 = \begin{cases} \text{'HL(+)} & \text{if } (\exists c \in e)(\exists r \in c.SY[])(r_{<disease_code>} \in \{\text{HL_DISEASE_CODES}\}) \text{ AND} \\ & (\exists c \in e)(\exists r \in c.IY[])(r_{<drug_code>} \in \{\text{HL_DRUG_CODES}\}) \\ \text{'HL(-)} & \text{otherwise} \end{cases}$

Parameters of $f_{tup.to.tup}$:

selection: $t \mid (t \in R)(t.V_3 = \text{'DM(+)} \text{ OR } t.V_4 = \text{'HBP(+)} \text{ OR } t.V_5 = \text{'HL(+)'})$

sql-based-conversion: **SELECT** $V_2, V_3, V_4, V_5, \text{COUNT}(V_1)$ **GROUP BY** V_2, V_3, V_4, V_5

図5 年度別の糖尿病・高血圧・高脂血症患者数の分析に於ける解析論理の記述例

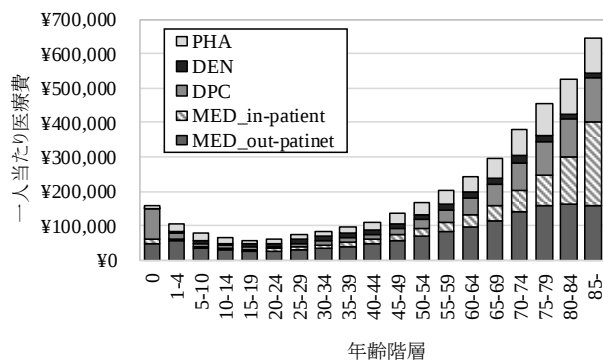


図4 特定年度に於ける年齢階層別の一人当たり医療費の分析の実行結果

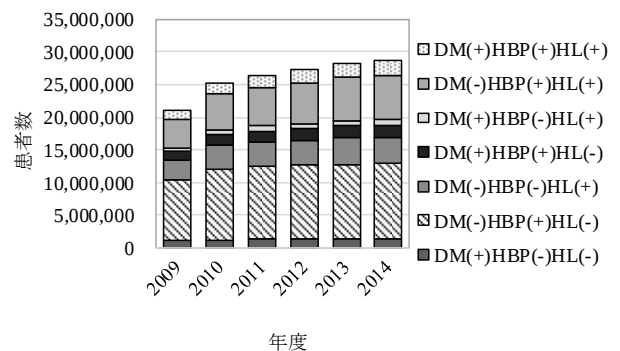


図6 年度別の糖尿病・高血圧・高脂血症患者数の分析の実行結果

Function flow definition: $input \xrightarrow{f_{rez.to.ep1}} E_1, input \xrightarrow{f_{rez.to.ep2}} E_2, \{E_1, E_2\} \xrightarrow{f_{eps.to.ep}} E_3 \xrightarrow{f_{ep.to.tup}} R_1 \xrightarrow{f_{tup.to.tup}} output$

Parameters of $f_{rez.to.ep1}$:

selection: $c \mid (c \in C)(c.RE_{<intervention_year_and_month>} \text{ in } [201304, 201309] \text{ AND } c.RE_{<age>} \geq 75)$ conversion: none
 sessionize: $key_1 = c.RE_{<patient_id>}$

Parameters of $f_{rez.to.ep2}$:

selection: $c \mid (c \in C)(c.RE_{<intervention_year_and_month>} \text{ in } [201310, 201609])$ conversion: none
 sessionize:

$$key_1 = c.RE_{<patient_id>}, key_2 = \begin{cases} '2013/10-' & \text{if } c.RE_{<intervention_year_and_month>} \text{ in } [201310, 201403] \\ '2014/04-' & \text{if } c.RE_{<intervention_year_and_month>} \text{ in } [201404, 201409] \\ '2014/10-' & \text{if } c.RE_{<intervention_year_and_month>} \text{ in } [201410, 201503] \\ '2015/04-' & \text{if } c.RE_{<intervention_year_and_month>} \text{ in } [201504, 201509] \\ '2015/10-' & \text{if } c.RE_{<intervention_year_and_month>} \text{ in } [201510, 201603] \\ '2016/04-' & \text{otherwise} \end{cases}$$

Parameters of $f_{eps.to.ep}$:

input: $E_1 = output(f_{rez.to.ep1}), E_2 = output(f_{rez.to.ep2})$ mergeKey: $E_1.e.c[0].RE_{<patient_id>} = E_2.e.c[0].RE_{<patient_id>}$

reduction:

$$E_1 : e \mid (\neg \exists c \in e)(\exists r \in c.SY[]) (r_{<disease_code>} \in \{DM_DISEASE_CODES\} \text{ AND } E_2 : \forall e$$

$$(\neg \exists c \in e)(\exists r \in c.IY[]) (r_{<drug_code>} \in \{DM_DRUG_CODES\}) \text{ AND}$$

$$(\neg \exists c \in e)(\exists r \in c.SI[]) (r_{<intervention_code>} \in \{DIALYSIS_CODES\})$$

Parameters of $f_{ep.to.tup}$:

selection: none conversion: none

summarization:

$t.V_1 = e.c[0].RE_{<patient_id>}$

$$t.V_2 = \begin{cases} '2013/10-' & \text{if } (\exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \text{ in } [201310, 201403]) \\ '2014/04-' & \text{if } (\exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \text{ in } [201404, 201409]) \\ '2014/10-' & \text{if } (\exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \text{ in } [201410, 201503]) \\ '2015/04-' & \text{if } (\exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \text{ in } [201504, 201509]) \\ '2015/10-' & \text{if } (\exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \text{ in } [201510, 201603]) \\ '2016/04-' & \text{otherwise} \end{cases}$$

$$t.V_3 = \begin{cases} 'NO-REZEPT' & \text{if } (\neg \exists c \in e)(\exists r \in c.RE)(r_{<intervention_year_and_month>} \geq 201310) \\ 'DM(+)DRG(+)DIA(+)' & \text{if } (\exists c \in e)(\exists r \in c.SY[]) (r_{<disease_code>} \in \{DM_DISEASE_CODES\} \\ & \text{AND } (\exists c \in e)(\exists r \in c.IY[]) (r_{<intervention_code>} \in \{DM_DRUG_CODES\}) \\ & \text{AND } (\exists c \in e)(\exists r \in c.SI[]) (r_{<drug_code>} \in \{DIALYSIS_CODES\}) \\ 'DM(+)DRG(+)DIA(-)' & \text{if } (\exists c \in e)(\exists r \in c.SY[]) (r_{<disease_code>} \in \{DM_DISEASE_CODES\}) \\ & \text{AND } (\exists c \in e)(\exists r \in c.IY[]) (r_{<intervention_code>} \in \{DM_DRUG_CODES\}) \\ & \text{AND } (\neg \exists c \in e)(\exists r \in c.SI[]) (r_{<drug_code>} \in \{DIALYSIS_CODES\}) \\ 'DM(+)DRG(-)DIA(+)' & \text{if } (\exists c \in e)(\exists r \in c.SY[]) (r_{<disease_code>} \in \{DM_DISEASE_CODES\}) \\ & \text{AND } (\neg \exists c \in e)(\exists r \in c.SI[]) (r_{<drug_code>} \in \{DIALYSIS_CODES\}) \\ & \text{AND } (\exists c \in e)(\exists r \in c.IY[]) (r_{<intervention_code>} \in \{DM_DRUG_CODES\}) \\ 'DM(+)DRG(-)DIA(-)' & \text{if } (\exists c \in e)(\exists r \in c.SY[]) (r_{<disease_code>} \in \{DM_DISEASE_CODES\}) \\ & \text{AND } (\neg \exists c \in e)(\exists r \in c.IY[]) (r_{<intervention_code>} \in \{DM_DRUG_CODES\}) \\ & \text{AND } (\neg \exists c \in e)(\exists r \in c.SI[]) (r_{<drug_code>} \in \{DIALYSIS_CODES\}) \\ 'DM(-)DRG(-)DIA(-)' & \text{otherwise} \end{cases}$$

Parameters of $f_{tup.to.tup}$:

selection: none sql-based-conversion: $\text{SELECT } V_2, V_3, \text{COUNT(DISTINCT } V_1) \text{ GROUP BY } V_2, V_3$

図7 非糖尿病患者群の重症化推移の分析に於ける解析論理の記述例

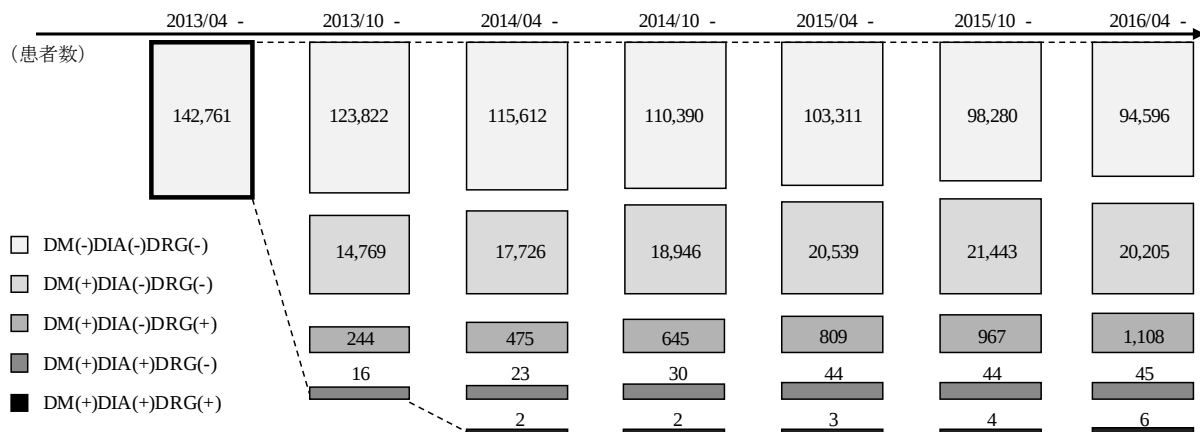


図8 非糖尿病患者群の重症化推移の分析の実行結果