

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

合田 和生[†] 山田 浩之[†] 喜連川 優[†] 満武 巨裕^{††}

[†] 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{††} 医療経済研究機構 〒105-0003 東京都港区西新橋 1-5-11 第 11 東洋海事ビル 2F

E-mail: [†]{kgoda,hiroyuki,kitsure}@tkl.iis.u-tokyo.ac.jp, ^{††}mitsutake@ihep.jp

あらまし 平成 21 年より、厚生労働省は、国内の公的医療保険を担う保険者等からの電子レセプト情報等の収集を継続的に実施しており、収集された情報を基にレセプト情報・特定健診等情報データベース（NDB）を構築し、医療費の急激な増大を抑制しつつ持続的な医療の提供を可能とするための医療政策立案への活用を試みている。著者らは、平成 24 年に、当該データベースから平成 22 年度 1 年分に相当する匿名化された大規模電子レセプト情報の提供を許可され、当該電子レセプト情報の高速な解析を可能とする実験システム「高速レセプト解析システム」を構築し、今日に至るまで医学分野の研究者等に解析サービスを提供している。本論文では、当該実験システムの概要ならびに高速な解析タスク処理を実現するためのデータベース構成を示すと共に、約 4 年に亘る開発・運用経験に基づき、将来に向けた課題と展望を纏める。

キーワード レセプト、公的医療保険、医療情報、システム開発、アウトオブオーダ型実行方式

1. はじめに

我が国の国民医療費^(注1)は、平成 26 年度に 40.8 兆円に達した。平成 6 年度には 25.8 兆円であったことから、この 20 年間で約 1.58 倍に増加している [8]。高齢化の進展、医療技術の高度化、疾病構造の変化等、多様な原因として推察されている [1, 3]。一方で、同じ 20 年間に国民総生産は 495.6 兆円から 489.6 兆円へと推移しており、年々の景況によって若干の変動は見られるものの、明確な成長の方向性は見られず、むしろ大局的には僅かな後退が見られる [15]。我が国が、その経済力に比して、医療費の国民負担が急増しており、今後も同様の傾向が続いた場合、現状の医療水準を国民が謳歌し続けることが不可能となるのは必至であろう。医療費の急激な増大を抑制しつつ、持続的な医療の提供を可能とすることは、我が国にとって極めて重要な国家的課題に他ならない。

我が国の医療は、社会保険料と公費（税金）を主たる財源としており、過去、高度な政策立案がなされてきた結果、他国と比べて遜色のない医療の効率性水準を達成している [2]。しかしながら、近年は公費による補填は徐々に増大してきており、平成 2 年に 25%であった公費補填率は、平成 22 年には 37%に膨らむに至っている [14]。給付と負担のバランス、即ち、社会保障制度の持続可能性は、実質的に損なわれており、抜本的な効率性の向上が欠かせない。国家レベルで医療の実態を徹底的に分析し尽くすことが、1 つの有力なアプローチと言え、多様な健康・医療分野のビッグデータの利活用への期待は高い。

著者らは、当該アプローチへの貢献を志し、厚生労働省から国内の 1 年分の全公的医療保険の請求書データである電子レセプト情報の提供を受け、当該情報の高速解析を可能とする実験システム「高速レセプト解析システム」を構築し、解析サービスを医療分野の研究者に対して提供してきた。本論文執筆時点で、運用実績は約 4 年に及び、徐々にその成果が明らかになりつつある。本論文では、「高速レセプト解析システム」の概要と特徴ならびに高速な電子レセプト解析を実現するために行ったシステム設計を明らかにすると共に、得られた解析事例を紹介する。当該システムは更に追加データの提供を受けて、拡充が予定されており、本論文では、将来に向けた課題と展望を纏める。

本論文の構成は以下の通りである。2. では、我が国の公的医療保険制度と、当該制度に於ける支払い実務を支えるレセプトならびにその電子化を説明する。3. では、「高速レセプト解析システム」の概要と特徴、解析事例ならびに高速なレセプト解析を実現するためのデータベース設計を明らかにする。4. では、これまでの経験に基づき、将来に向けた課題と展望を示し、5. に於いて本論文を纏める。

2. 我が国の公的医療保険制度と電子レセプト情報

2.1 我が国の公的医療保険制度

医療保険は、被保険者が医療機関を受診し、医療従事者から診療、処置、投薬等の医療行為を受けることによって発生した対価である医療費について、その一部もしくは全部を保険者が給付する制度であり、長期の入院や先端的な治療を受けることにより生じる高額な医療費を、被保険者が直接的に負担することを避けることにより、被保険者の健康と医療機関の提供す

(注1)：保険診療の傷病の治療に要した費用の合計を意味する。例えば、評価療養、不妊治療における生殖補助医療、正常な妊娠・分娩、健康の維持・増進を目的とした健康診断や予防接種等に要する費用等は除外される。

る医療サービスを高い水準で維持することを目的とする。医療保険には多様な仕組みが存在するが、行政等が予め被保険者の範囲を指定して被保険者を強制的に加入させる公的医療保険と、任意加入の私的医療保険の2種類に大別することができ、我が国を含む多くの先進国では、公的医療保険と私的医療保険が併用されている。

とりわけ我が国に於いては、昭和36年以降、政策として所謂、**国民皆保険制度**が採用され、全国民^(注2)は、何らかの公的医療保険に加入することを義務付けられており、公的医療保険が我が国の医療制度全般を支える礎としての役割を担っている。我が国の主な公的医療保険を表1に纏める。他方、私的医療保険は、我が国では、公的医療保険を補完するものとして位置付けられており、公的医療保険の自己負担分の補填や休職時の収入の補償等に留まっているのが実状である。

2.2 レセプトとその電子化

公的医療保険が適用される保険診療に於いては、医療機関(病院、診療所、調剤薬局等)は、毎月、保険者に対する請求書を予め決められた様式で作成し、保険者に送付する。また、これを受領した保険者は、請求された医療費を審査し、医療費のうち保険者負担分を医療機関に支払う。この際の請求書は、**レセプト**(診療報酬明細書; medical insurance claim; (独)Rezept)と呼ばれ、請求元である医療機関の情報に加えて、被保険者毎に、患者の個人情報、傷病名、提供した医療行為、診療報酬ならびに参考情報等が纏められて、記載される。我が国では、医療行為の対価である診療報酬は、出来高払いを原則としており、医療行為毎に対価が定められており、提供した医療行為に対して対価を合算して請求する。ただし、一部の病院に於ける急性期入院については、疾病群別包括払い制度(diagnosis procedure combination; DPC)が導入されており、当該制度に当たる場合は、医療行為ではなく臨床疾患エピソードによって診療報酬が決定される。医療機関の種別等に応じて、記載様式が定められており、レセプトの様式は主に以下の4つに分類することができる。

- **医科用レセプト**：病院や診療所に於いて患者が外来診療もしくは入院診療を受けた際に発生する
- **DPC用レセプト**：特定機能病院等に於いて患者が疾病群別包括払い制度が適用された入院診療を受けた場合に発生する
- **歯科用レセプト**：病院や診療所に於いて患者が歯科診療を受けた際に発生する
- **調剤用レセプト**：調剤薬局に於いて患者が調剤を受けた際に発生する

従前は紙媒体によってレセプトを提出することが手続きとして求められていたが、ようやく平成11年に磁気媒体等による電子的な手段による提出が認められて以降、当該機能を備えたレセコンと呼ばれる端末やソフトウェアが医療機関に普及し、平成27年4月診療分からは原則、電子レセプトによる請求が

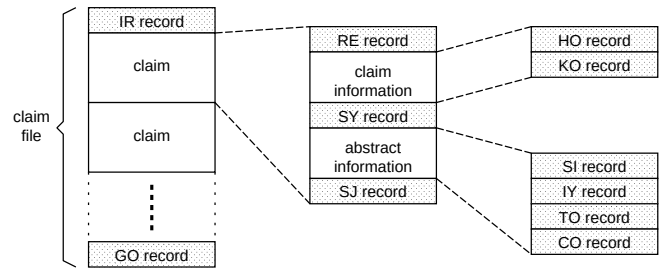


図1 電子レセプト情報を格納するファイルの編成の一例(医科用レセプトの場合)^(注5)

原則化されており、平成28年10月期では、電子化率は医療機関数ベースで92.5%、請求件数ベースで98.0%に達している[12]。この際、レセプトは、任意の医療機関と任意の保険者の間で交換可能であることが不可欠であることから、電子化に際して、従来の紙媒体の様式を基に、電子的な様式の仕様が定められた。その後、政策や規則の変化に伴い、レセプトの様式仕様は改訂が繰り返されている[9]。

当該仕様によれば、電子レセプト情報を格納するファイルは所謂CSV(comma-separated values)形式のテキストで表現され、テキスト中の各行はレコードと呼ばれ、1番目のカラム(英数字2文字の文字列)がその種別を表す識別子である。ファイル編成の一例を図1に図示すると共に、また、電子レセプト情報を構成するレコードの種別の一例を表2に纏める。このように、電子レセプト情報は、入れ子タプル(nested tuple)構造を有しており、REレコードを先頭とする一連のレコードが、1件の請求を構成する。即ち、1件の請求は、1件のREレコードから開始し、次のREレコードの直前で終了する^(注3)。この際、ある種のレコード(例えば、傷病名を表すSYレコード)は、1件の請求に複数が含まれることがあり、また、ある種のレコード(例えば、公費負担を表すKOレコード)は、含まれないこともある。図2に、電子レセプトファイルの一例(医科用レセプト)を示す。ここでは、2件の請求が含まれている。

電子レセプト情報に含まれる情報のうち、傷病名、診療行為、医薬品、特定機材等の情報については、別にマスタ仕様が定められている[9]。例えば、SYレコードの2番目のカラムは、患者の傷病名を記載する項目であるが、予め定められた7桁の数値によって傷病名を記載する必要がある。例えば、図2の4行めと15-16行めのSYレコードは、それぞれ、「早期胃癌ESD後」(8845988)、「扁桃肥大」(8840014)、「急性中耳炎」(3829027)を意味する。なお、従来に紙媒体で記入されていた慣習や、臨床現場の医師の判断により、マスタ情報に含まれていない値の記入を要する場合があることから、例外的な値を記入することが認められている。例えば、傷病名に関しては、患者が傷病名マスタに登録されていない傷病名の疑いがあること

(注2)：厳密には、生活保護の受給者等を除く国内に居住する全国民と一定の在留資格のある外国人を、被保険者とする。

(注3)：実際の電子レセプト情報には、手続きの過程で、更に審査状況等を表す別のレコードが付加されて処理されるが、本論文ではこれを議論しない。

(注5)：本論文の議論を越えることから、国保連固有情報レコード(KH)、包括評価対象外理由レコード(GR)、臓器提供者関連レコード(TI, TR, TS)は簡単のために省いている。

表 1 我が国の主な公的医療保険

制度の種別	保険制度	保険者	被保険者
被用者保険	協会管掌健康保険	全国健康保険協会	健康保険組合を組織しない企業の従業員とその被扶養者
	組合管掌健康保険	健康保険組合	健康保険組合を組織する企業の従業員とその被扶養者
	共済管掌健康保険	共済組合	公務員や私立学校教職員等とその被扶養者
	船員保険	全国健康保険協会	船員とその被扶養者
地域保険	国民健康保険	市町村と東京都特別区、国民健康保険組合	個人事業主、無職者等
	後期高齢者医療制度	後期高齢者医療広域連合	後期高齢者（75 歳以上の者）等

表 2 電子レセプト情報を構成する主なレコードの種別（医科用レセプトの場合）

種別	識別子	レコードの意味
医療機関情報	IR	保険医療機関に関する情報（医療機関コード、医療機関名称、請求年月等）を表す。医療機関単位のデータの先頭を意味する。
レセプト共通	RE	レセプトの基本情報（診療年月、レセプト種別、患者の氏名等の個人情報等）を記録する。患者単位のデータの先頭を意味する。
保険者	HO	医療費の請求先保険者に関する情報（保険者番号、被保険者番号、診療実日数、合計点数等）を記録する
公費	KO	医療費の公費負担に関する情報（負担者番号、受給者番号、診療実日数、合計点数等）を記録する
傷病名	SY	患者の傷病名に関する情報（傷病名コード、診察開始日、転帰区分等）を記録する
診療行為	SI	患者に行った診療行為に関する情報（診療行為コード、数量データ、点数、回数等）を記録する
医薬品	IY	患者に使用した医薬品に関する情報（医薬品コード、使用量、点数、回数等）を記録する
特定機材	TO	患者に使用した特定機材に関する情報（特定機材コード、使用額、点数、回数等）を記録する
コメント	CO	請求に関する医療機関のコメント（コメントコード、文字データ等）を記録する
症状詳記	SJ	患者の症状の詳細情報（症状詳記区分、症状詳記データ等）を記録する
診療報酬請求	GO	保険医療機関の診療報酬請求の合算情報（総件数、総合計点数等）を記録する。医療機関単位のデータの末尾を意味する。

を意味するために、未コード化傷病名（0000999）なる専用の値が定められている。

2.3 レセプト情報・特定健診等情報データベース

レセプトは、飽くまで医療費の請求を元来の目的とするものであるから、検査結果値や画像等の臨床情報は含まれないものの、医療従事者が実施した医療行為に関する情報等は含まれている。既に述べたように、我が国は国民皆保険制度を採用しており、極一部の例外的なケースを除き、国内で提供される傷病の治療を目的とするほぼ全ての医療は、公的医療保険によって提供がなされていると言っても過言ではなく、その請求情報は統一的な仕様によって電子化されている。即ち、我が国で生成

```

1  IR,1,13,1,9999939,,サンプル医科病院,42811,00,03-9999-9999
2  RE,3,1118,42805,サンプル医科0
3  2,1,3180717,,,,,500,,,,,,01,,,,,,
4  HO,06132013,1 2 3 4 5 6 7,1,1,141,,,,,
5  SY,8845988,4280527,1,,,01,
6  CO,01,1,810000001,他科受診中にて初診料算定せず
7  SI,12,1,112011310,,73,1,,,,,,1
8  SI,80,1,120002910,,68,1,,,,,,1
9  CO,,1,810000001,内科
10 IR,1,13,1,9999939,,サンプル医科病院,42811,00,03-9999-9999
11 RE,9,1124,42805,サンプル医科0
12 6,1,4240223,,,,,500,,,,,,27,,,,,,
13 HO,06132013,1 2 3 4 5 6 7,1,2,222,,,,,
14 KO,88132121,2222222,,2,222,,,,,
15 SY,8840014,4271216,1,,,01,
16 SY,3829027,4271216,1,,,
17 SY,4742002,4271216,1,,,
18 SY,4739014,4271224,1,,,
19 SY,3814010,4280106,1,2058,,
20 SY,8832434,4280323,1,,,
21 SY,0389004,4280427,1,,,
22 SY,4640013,4280427,1,,,
SI,12,2,112011310,,2,,,,,,1,,,,,1,,
SI,,2,112006270,,111,2,,,,,,1,,,,,1,

```

図 2 電子レセプト情報の一例（医科用レセプト、2 件の請求分）

される電子レセプト情報を一元的に収集し、解析することは、国内で提供されている医療を網羅的に把握し、国家の医療政策の立案等に活用するために、極めて有力なアプローチと考えられる。

平成 20 年に改正された「高齢者の医療の確保に関する法律」（昭和 57 年法律第 80 号）に於いては、第 16 条（医療費適正化計画の作成等のための調査及び分析等）によって、厚生労働省に保険者等から必要な情報を収集して解析する権限が与えられた。翌年から今日に至るまで、同省は保険者等から匿名化された電子レセプト情報等の継続的な収集を実施しており、収集された情報により構築された**レセプト情報・特定健診等情報データベース**（NDB^(注6)）[13, 16] の規模は、平成 28 年度末には請求約 100 億件、レコード約 2,600 億件に達するものを推測される。国家レベルの悉皆性を有する健康・医療分野のビッグデータは世界的にも稀有であり、当該データベースは其中でも随一の規模を誇り、我が国の医療水準の向上や医療費の抑制に向けた政策立案への活用が期待されている^(注7)。

(注6)：国家レベルの悉皆性を備えたデータベースを意味して、ナショナルデータベース（national database; NDB）と通称されることが多い。

(注7)：レセプト情報・特定健診等情報データベースには、電子レセプト情報に加えて、特定検診・特定保険指導データも格納されており、その利活用への期待は高い。紙面の制約上、当該議論は別稿に譲る。

表 3 提供を受けた平成 22 年度分の電子レセプト情報の諸元

レセプト種別	請求件数	レコード数
医科（外来分）	888,368,955 件	18,907,146,341 件
医科（入院分）	17,482,002 件	385,860,129 件
DPC	10,672,681 件	2,099,584,737 件
歯科	31,676,839 件	485,780,189 件
調剤	562,836,049 件	8,582,030,136 件
合計	1,511,036,526 件	30,460,401,532 件

3. 高速レセプト解析システム

3.1 電子レセプト情報の第三者提供

レセプト情報・特定健診等情報データベースは、第一義的には、厚生労働省ならびに都道府県等の行政をユーザとし、医療費適正化のための分析やエビデンスの抽出を目的としているものの、第三者提供と称する枠組みが設けられており、公益性の高い研究に対しては、「レセプト情報等の提供に関する有識者会議」の承諾の下で、その一部のデータを加工して、提供することが認められている。当然の事ながら、高いセキュリティ基準を満たす事等、提供を受けるための要件を満たすことは容易ではない。

著者らは、平成 24 年に、当該データベースから平成 22 年度 1 年分の電子レセプト情報を提供をお願いする申請を行い、有識者会議の承諾を得て、厚生労働省から提供を受けることとなった。提供を受けるに当たり、被保険者の特定をより困難化するために匿名化レベルを上げる、医療機関コードをハッシュ関数に掛けて医療機関を容易に特定できないようにする、医薬品コードを製品名を容易に特定できないように一般名を値としたものに変換する等の幾つかの加工が行われたものの、1 年分の全ての電子レセプト情報が提供されることとなり、潜在的に、国内で提供されたほぼ全ての医療を網羅的に解析する機会を得たと言える。表 3 に提供を受けた電子レセプト情報の諸元を纏める。テキストファイルとしての容量は 2.2TB であった。

3.2 高速レセプト解析システムの構築と医療分野の研究者に対する解析サービスの提供

従前より、医学分野の研究者は、通常、スプレッドシートや統計解析等のソフトウェアツールを用いて、電子レセプト情報の分析を行っていたが、一般にこれらのツールは、高々主記憶に容易に収まる規模のデータを処理するように設計がなされており、1 年分の我が国の全電子レセプト情報を機動的に分析するには、明らかに非力と言え、本格的な高速データベースシステムを用いた解析システムの構築が欠かせない。他方、当時の厚生労働省は、レセプト情報・特定健診等情報データベースを管理するための大規模システムを保有していたが、当該システムは、制度が開始されて間もない頃に設計されたものであり、電子レセプト情報等の継続的な収集に重きが置かれ、解析については十分に検討がなされたとは言えない状況であったことから、収集された大規模電子レセプト情報の本格的な活用には至っていなかった。

著者らは、厚生労働省から大規模電子レセプト情報の提供を

受けた際に、情報学分野の研究者と医学分野の研究者が協働して、当該大規模情報の解析システムのあり方を探求し、その成果を学術界や行政関係者に還元することは意義深いものと考え、東京大学で開発した**アウトオブオーダ型高速データベースエンジン** [6] を核とする実験システム「高速レセプト解析システム」[11] を、平成 24 年に東京大学生産技術研究所内に構築し、当該実験システムを用いて、医学分野の研究者に対する解析サービスの試験的な提供を開始した。

当該実験システムの主な特徴を以下に纏める。

- 医療分野の研究者が、容易かつ機動的に解析を進めることが出来るように、図 3 に一例を示す如く、解析のシナリオに応じた直感的な Web インターフェースを開発した。当該インターフェース上では、医学分野の用語で解析タスクを設定できるものとし、解析結果はスプレッドシート、グラフ、地図等で表示され、研究者が即座にこれを確認して、次の解析タスクを検討できるようにした。また、解析シナリオによっては、試行錯誤のためのドリルダウン操作が多発するため、グラフや地図上でのポインティングデバイス操作により、直ちにドリルダウン用の解析タスクを駆動できるようにした。

- 電子レセプト情報には、傷病名や医薬品には、独自の識別用コードが規格化され、当該コードを用いて記録されている。他方、医学分野では、これとは異なる仕組みで、傷病名や医薬品の体系化（ICD-10 や H0T-9 等）が進んでいることから、データベース内でマスタ情報を用いてコード変換を行うことにより、医学分野で体系化されたコードを用いた解析タスクの設定を可能とした。

- 主にレセプトには、医科、DPC、歯科、調剤と異なる種類が存在し、様式が異なるが、異なる種別のレセプトを結合し、患者単位で集約して扱うことにより、レセプト種別を横断した統一的な解析タスクの設計と実行を可能とした。

- 全ての解析タスクと解析結果をログに記録することにより、医療分野の研究者の利用動向を把握可能とし、基盤の改良に活用できるようにした。

- 実験システムは、表 4 に纏めるハードウェア上で構築した。医療分野の研究者による利用が著しい場合には、最大で常用サブシステムよ拡張サブシステムの全てのノードを当該実験システムに割り当てて用い、他方、実験システムの利用が著しくない場合には、常用サブシステムのみを割り当てることとした。

当初は、医学分野の研究者からは、極めて限定的な解析タスクのみの実行を想定しているとの話を聞かされていた。しかしながら、実際に実験システムが稼働するようになると、想定より機動的に解析が進むことから、新たなタイプの解析タスクを実行したいとの要望が、次々と著者らに寄せられるようになり、これらの要望に答えるべく、実験システムは断続的に進化することとなった。現在に於いても、当該システムは、新たな解析タスクを実行可能とするべく、1 月～2 月に 1 回の頻度で機能追加を続けている。

3.3 電子レセプト情報の解析事例

前節で述べた「高速レセプト解析システム」を用いて得られ

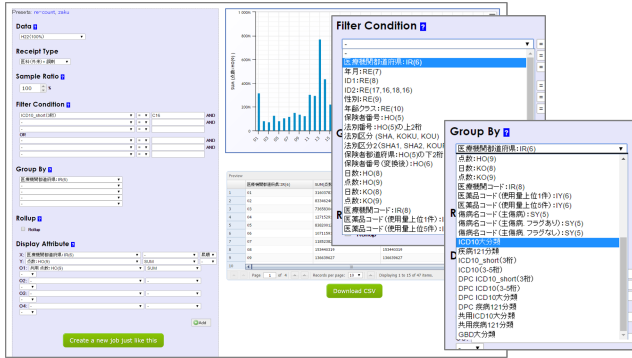


図 3 実験システムの Web インターフェースの一例

表 4 実験システムのハードウェアの諸元

常用サブシステム	
処理ノード (1 ノード)	Intel Xeon E7-2870 2.4GHz (2p/20c), 512GB Memory, 16x 7.2Krpm 3TB HDDs, 1Gbps NIC, CentOS Linux
拡張サブシステム	
処理ノード (128 ノード)	Intel Xeon E5-2680 2.7GHz (2p/16c), 64GB Memory, 24x 10Krpm 900GB SAS HDDs, 2x 15Krpm 300GB SAS HDDs, 1Gbps/10Gbps NICs, Cen- tOS Linux (ノード当たりの諸元)
データ用スイッチ	1x 128 ports 10Gbps
制御用スイッチ	6x 48 ports 1Gbps

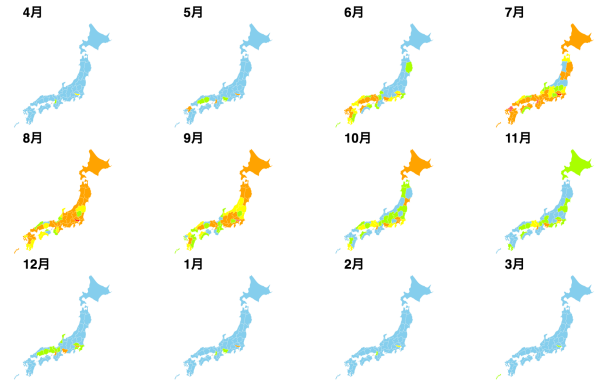
た解析結果の一例を示す。

図 4(a) は、各月・各都道府県に於ける手足口病なる傷病の発生率（患者数を人口で割ったもの）を表したものである。青色は発生率が極めて低いことを表し、赤色に近づく程、発生率が高いことを意味する。

図 4(b) は、各都道府県別に、1 年間に於ける CT (Computed Tomography; コンピュータ断層撮影法), MRI (Magnetic Resonance Imaging; 磁気共鳴画像), PET (Positron Emission Tomography; 陽電子放出断層撮影法) の各検査機器の利用率（利用件数を人口で割ったもの）を表したものである。同様に、青色は利用率が極めて低いことを表し、赤色に近づく程、利用率が高いことを意味する。

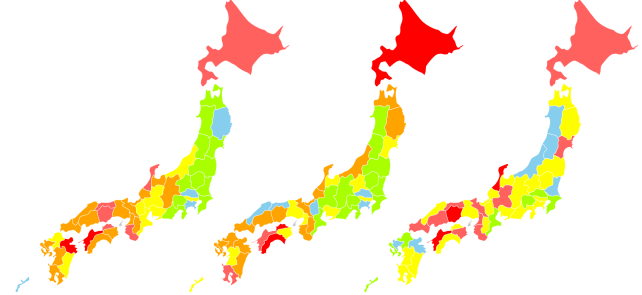
このように、「高速レセプト解析システム」では、例えば、傷病や医療機器等の属性を指定して、我が国の網羅的な医療動向をインタラクティブに表示することを実現している。従前の調査が、特定日のみを対象とした抽出調査に留まっていたのに対して [7], 次々と視点を変えて、全数を対象として解析を進めることが出来る点が、得に医学分野の研究者ユーザから好評を得ている [17, 18]。

なお、上記は、厚生労働省の運用するレセプト情報・特定健診等情報データベースから提供を受けた電子レセプト情報の解析事例であるが、最近では、地方自治体から直接、夫々が管掌する医療等社会保険（国民健康保険、後期高齢者医療制度、介護保険等）に関する電子レセプト情報が著者らに提供されるようになってきており、「高速レセプト解析システム」による解析



(a) 手足口病の月別・都道府県別の発生率の概況

CT の利用件数の対人口比率 MRI の利用件数の対人口比率 PET の利用件数の対人口比率



(b) 高額医療機器 (CT, MRI, PET) の利用率の概況

図 4 電子レセプト情報の解析事例 (国内全数)

サービスを研究者や一部行政関係者に提供している。地方自治体から提供される電子レセプト情報は、レセプト情報・特定健診等情報データベースから提供を受けているものと原理的には同一であるが、地方自治体のニーズに基づき、レセプト情報・特定健診等情報データベースからは提供されない情報や当該データベースが保有しない情報も提供されるようになってきており、より緻密な解析を可能としている点が特徴的である。図 5(a)(b) は、三重県に於ける国民健康保険等被保険者のうちの胃癌患者の所在地と、通院先医療機関の別に、患者の数を可視化したものである。丸の面積が患者数を表し、赤色が所在地別、青色が通院先医療機関別を意味する。地域によって様々である患者の通院動態を、的確に把握することが出来るツールとして、とりわけ行政関係者から高くご評価を頂いている。

3.4 データベースの設計とアウトオブオーダ型実行方式に基づく解析タスクの実行

既に述べた如く、電子レセプト情報は、入れ子タプル構造を有することから、今日に於いて一般的なデータベースシステムが対象とする関係データモデルを、そのまま適用することはできない。データベース構造を設計するアプローチとしては、以下の 2 つが考えられる。

- 関係データモデルを直接適用することが可能なようにデータ構造を変換して、既存の関係データベースシステムを用いて電子レセプト情報を格納して管理する。例えば、図 6 に図示する様に、レセプト種別毎に関係表を作成し、外部鍵参照を用いてレセプト種別間のレコードの関係性を表現するといったスキーマ構造が考えられる。

- 電子レセプト情報をそのまま格納して管理できるように、

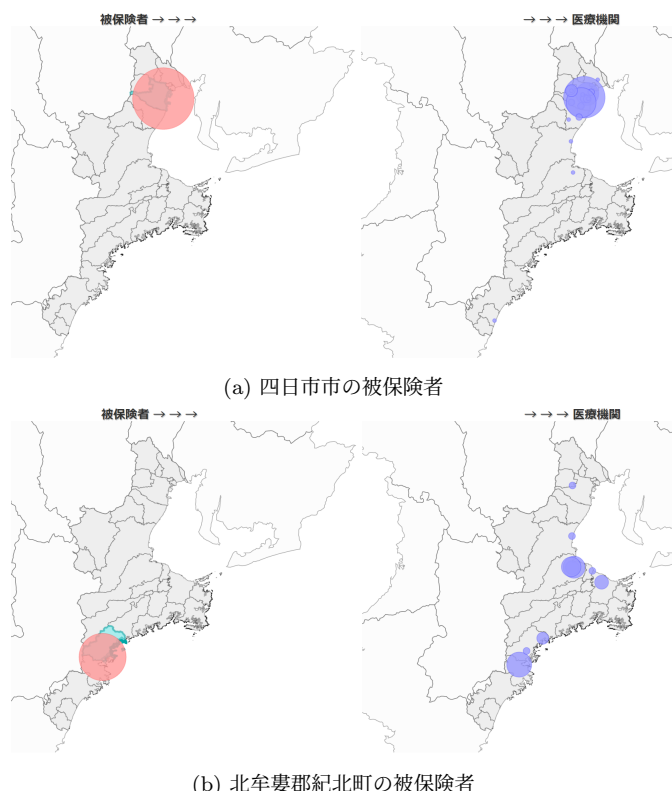


図 5 胃癌患者の所在地・通院先医療機関別の患者数の可視化事例（三重県国民健康保険被保険者等）

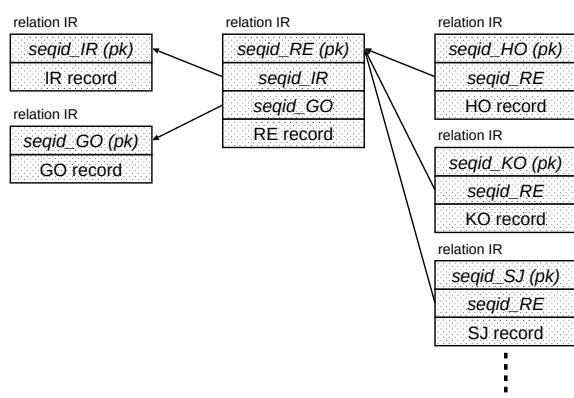


図 6 電子レセプト情報を関係データベースに格納する場合のスキーマ構造の一例

データベースシステムに新たな機能を設ける。

「高速レセプト解析システム」を構築するに当たり、上記 2 つのアプローチを検討したところ、一般に産業界での実装を想定すると、既存のデータベースシステムをそのまま用いることが出来ることから、前者のアプローチが有力であると考えられるものの、大学に於ける学術研究としては、より柔軟性の高い後者のアプローチによってシステムを構築した上で、当該システム上で模擬的に前者のアプローチの如きスキーマ構造を組み込むことにより、両者のアプローチの有用性を比較することが可能であると見込まれることから、後者のアプローチに従って実験システムを構築することにより、より広汎かつ深度の深い知見が得られる可能性が高いと判断し、これを採用することと

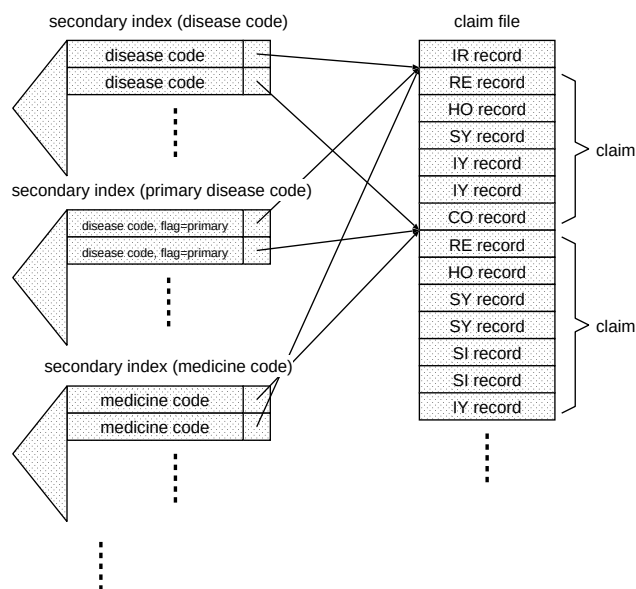


図 7 データベースの設計

した。

当該アプローチに従って、著者のうち合田と山田が設計した電子レセプト情報のためのデータベース構造を図 7 に図示する。電子レセプト情報は、レセプト種別毎に単一の論理ファイルとして、二次記憶空間に格納する。この際、二次索引等の補助データ構造を構築し、効率的に論理ファイル中のレセプトレコードにアクセスできるようにする。例えば、単一の請求を表す一連のレセプトレコード中には、複数の SY レコードが含まれる可能性があり、即ち、複数の傷病名が記録される場合があるが、二次索引中のこの傷病名コードに対するエントリ毎に、元の請求を表す一連のレセプトレコードへの参照を保持することにより、傷病名を元としたレセプトの効率的な検索を可能とした。また、SY レコードには、記録された傷病名が主傷病名^(注8)であるかどうかのフラグが存在するが、主傷病名のみによる効率的な検索を可能とするために、複合性を有する索引等、柔軟な索引構築を可能とした。なお、補助データ構造の構築とそれを用いた検索の機能は、文献 [10] に示す構造定義方式に従って実装し、アウトオブオーダ型データベースエンジンにプラグインした。

上述の設計によるデータベースに対して、表 5 に示す試験解析タスクを実行し、その実行時間を計測した。この際、実験システムのハードウェア構成としては、表 4 に示す拡張サブシステムの全 128 ノードを用いることとし、アウトオブオーダ型データベースエンジンの実装としては、並列データ処理系 Hadoop にアウトオブオーダ型実行方式を適用した実装 [10,11] を用いた。アウトオブオーダ型実行方式の有効性を確認するために、比較対象として、並列データ処理系 Hadoop に於いても同じ試験解析タスクを実行し、その計測結果を比較することとした。計測結果を図 8 に図示する。アウトオブオーダ型実行方

(注8)：主傷病名は医学上明確に定義されている訳ではないが、レセプトに記録した傷病名のうち 1 つを主表病名に指定することが求められている。

表 5 試験用の解析タスク

番号	解析タスクの出力
Q.1	急性心膜炎の患者の数の都道府県毎の対人口比率
Q.2	口唇の悪性新生物の患者に掛かる都道府県毎の平均診療費
Q.3	呼吸器結核の患者の年齢毎のレセプト件数
Q.4	コレラの患者に掛かる都道府県毎の平均診療費



図 8 性能試験の結果

式を用いることにより、93 倍から 117 倍程度的高速化が達成され、試験解析タスクに関して概ね 1 秒程度の応答性を実現できることを確認した。

先に述べたように、解析シナリオによっては、試行錯誤のためのドリルダウン操作が多発することから、当該プロセスの機動性を高めることが課題であった。提案する実験システムは、少なくとも我が国の 1 年分の全電子レセプト情報に関して、選択性のある解析タスクについては、1 秒程度の応答性を実現できており、その有用性は著しいと言える。

4. 課題と展望

著者らが実験システム「高速レセプト解析システム」の構築を開始してから、本論文の執筆時点で約 4 年が経過した。その間、情報学分野の研究者と医学分野の研究者との協働により、機動的かつ断続的に機能を追加し、着実に実験システムを成長させて来た。当該成果が認められ、本年度には、新たに、レセプト情報・特定健診等情報データベースから 5 年分（平成 21 年度分、平成 23～26 年度分）の全電子レセプト情報の著者らの研究グループへの第三者提供が、有識者会議に於いて承諾され、現在、データの受渡し作業が進んでいる。今後数ヶ月以内に、6 年分の我が国を網羅した電子レセプト情報を解析可能とすべく、実験システムの拡充を進めているところである。本節では、以降、主に情報学分野の視点から、今後の研究に向けた課題と展望を以下に纏める。

a) 実際のビッグデータに対するシステム技術の追求

情報システムの学術研究に於いて、システム技術を考案し、検証するに当たり、実際に意味を有し、これ程の規模のデータセットを対象とする機会は、必ずしも多くはなく、貴重と言える。殊に、本実験システムに於いては、データの規模が現状の 1 年分から 6 年分に急増し、少なくともテキスト情報として 15TB 程度に達するものと推察される。また、この後、詳細

を記載する如く、解析タスクの複雑度が向上し、ユーザ数が拡大するのは必至と言える。医学分野の研究者の解析業務を促進すべく、応答性の高い解析タスクの処理と、機動的なソフトウェア機能の拡充とを可能とする効率的なシステムを如何に実現するかは、システム技術の研究にとって良き実問題と言える。また、既に、レセプトは、診療報酬の単なる請求書としての役割を越え、医療政策の立案に不可欠なエビデンス収集のための観測網としての機能がむしろ広く認識されるようになりつつあり、後年の解析への活用を想定して、従来は無かった診療日付等の項目が次々と追加されるようになってきている。既に蓄積された大規模電子レセプト情報を如何に活用するかという现阶段の問題と並行して、今後如何に更なる大規模電子レセプト情報を収集し、これを解析していくかという将来に向けたグランドデザインが、国家戦略上、重要な課題と言える。東京大学で構築し、運用している「高速レセプト解析システム」は、当該デザインを描く上で、又とない実験場と言え、研究成果を還元できるものと期待している。

b) より深度の高い解析の実現

これまでに東京大学の実験システムに於いて、医学分野の研究者が希望してきた解析タスクは、ある期間内に於いて、単一もしくは複数の条件（例えば、特定の質病名を抱えている、特定の医薬品を処方されている等）を満たす被保険者群について、医療費や診療日数等の統計量を算出する類のものが典型例であった。レセプトは、1ヶ月単位で生成され、そこには検査値等の臨床情報が記録されないことから、患者の臨床像に於ける急峻な変化を正確に捉えることは困難性が高く、むしろ、網羅的に患者や医療期間の実態を把握することが、優先的に試みられてきたのは自然な流れと言える。しかしながら、6 年分の電子レセプト情報の解析が可能となると、患者の臨床像に於ける緩やかな変化を、部分的にでも捉えられる可能性が生じるものと推察される。主記憶に収まらない規模のデータに対する効率的な時系列解析アルゴリズムや、医学分野の研究者がこれを駆使することを可能とする高度な可視化インターフェースが活躍する機会が大きいものと考えられる。とりわけ、我が国で逼迫する医療費の多くは、生活習慣病がその根源にあると言われていた [4,5]。当該疾患は長年を掛けて形勢されるものであるから、健康状態から疾患へと至る時系列的な変化をこれまで把握する術はなかったが、国家レベルの悉皆分析により、その一端でも解明することが出来れば、インパクトは絶大なものと期待している。

c) システムのユーザ層の拡大

ユーザ層の拡大も重要な課題と言える。現状では、レセプト情報・特定健診等情報データベースから第三者提供されたデータセットは、有識者会議から承認を受けた研究者のみが利用可能であり、個人情報に準じる扱いを行うために、研究者は所属機関に於いて倫理審査による承認を得る必要がある。また、東京大学の構内に設けた専用のセキュリティルームに於いてのみ、アクセスが認められていることから、現実には実験システムを利用可能な研究者は若干名に限られているのが実状である。研究領域の細分化は進んでいる現状を鑑みると、我が国の電子レセ

プト情報の分析を推進し、医療水準の向上と医療費の抑制に繋げるためには、ユーザ層の拡大が欠かせないことは明らかであるが、現状は必ずしもそのために適した状態とは言えない。当然の事ながら、利便性と規制のバランスには、国民的な合意形勢が不可欠であり、社会的な便益に繋がる成果を如何に創出するかが鍵と言え、著者らは構築した「高速レセプト解析システム」が貢献できる潜在性は高いと言える。

5. おわりに

著者らは、厚生労働省が構築したレセプト情報・特定健診等情報データベース（NDB）から、平成22年度1年分に相当する匿名化された大規模電子レセプト情報の提供を受ける機会を得て、平成24年に、当該ビッグデータを解析するための実験システム「高速レセプト解析システム」を構築した。今日に至るまで、継続的に機能を追加すると共に、運用を行い、医学分野の研究者等に解析サービスを提供している。本論文では、当該実験システムの概要ならびに高速な解析タスク処理を実現するためのデータベース構成を示すと共に、約4年に亘る開発・運用経験に基づき、将来に向けた課題と展望を纏めた。

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」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。

付 録

本研究の一部は、東京大学ライフサイエンス委員会倫理審査専門委員会（審査番号：14-20, 15-36）による承認の下に行われた。

文 献

- [1] Victor R. Fuchs. The Gross Domestic Product and Health Care Spending. *New England Journal of Medicine*, Vol. 369, No. 2, pp. 107–109, 2013.
- [2] Peter Groves, Basel Kayyali, David Knot, and Steve van Kuiken. The 'big data' revolution in healthcare: Accelerating value and innovation. McKinsey & Company, 2013.
- [3] Sheila Smith, Joseph P. Newhouse, and Mark S. Freeland. Income, Insurance, And Technology: Why Does Health Spending Outpace Economic Growth? *Health Affairs*, Vol. 28, No. 5, pp. 1276–1284, 2009.
- [4] 伊藤由希子, 川渕孝一. 生活習慣病予防事業が医療費に及ぼす効

- 果：トヨタ自動車健康保険組合データを用いた検証. 東京学芸大学紀要 人文社会科学系 II, Vol. 61, pp. 155–171, 2010.
- [5] 岡村智教, 中村幸志, 早川岳人, 神田秀幸, 三浦克之, 岡山明, 上島弘嗣. 生活習慣病の予防と医療費：10年間の追跡調査による健診所見と医療費の関連：滋賀国保コホート研究の知見から. *日本衛生学雑誌*, Vol. 67, No. 1, pp. 38–43, 2012.
 - [6] 喜連川優, 合田和生. アウトオブオーダーデータベースエンジン OoODE の構想と初期実験. *日本データベース学会論文誌*, Vol. 8, No. 1, pp. 131–136, 2009.
 - [7] 厚生労働省. 厚生労働統計：患者調査. <http://www.mhlw.go.jp/toukei/list/10-20.html>. 2017年1月15日に参照.
 - [8] 厚生労働省. 厚生労働統計：国民医療費. <http://www.mhlw.go.jp/toukei/list/37-21.html>. 2017年1月15日に参照.
 - [9] 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/>. 2017年1月15日に参照.
 - [10] 山田浩之, 合田和生, 喜連川優. Hadoop をはじめとする並列データ処理系へのアウトオブオーダー型実行方式の適用とその有効性の検証. *電子情報通信学会論文誌*, Vol. J97-D, No. 4, pp. 774–792, 2014.
 - [11] 山田浩之, 合田和生, 喜連川優. 128 ノード規模のストレージインテンシブクラスタ環境におけるアウトオブオーダー型並列データ処理系の性能評価. *電子情報通信学会論文誌*, Vol. J98-D, No. 5, pp. 728–741, 2015.
 - [12] 社会保険診療報酬支払基金. レセプト請求形態別の請求状況（平成28年度）：平成28年10月診療分. http://www.ssk.or.jp/tokeijoho/tokeijoho_rezept/tokeijoho_04_h28.files/seikyu_2810.pdf. 2017年1月15日に参照.
 - [13] 藤森研司. レセプトデータベース（NDB）の現状とその活用に対する課題. *医療と社会*, Vol. 26, No. 1, pp. 15–24, 2016.
 - [14] 内閣府. 経済・財政一体改革推進委員会, 社会保障ワーキング・グループ資料. <http://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/wg1/281013/shiryuu1-2.pdf>. 2017年2月17日に参照.
 - [15] 内閣府. 国民経済計算（GDP 統計）. <http://www.esri.cao.go.jp/jp/sna/menu.html>. 2017年1月15日に参照.
 - [16] 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*, Vol. 56, No. 2, pp. 140–144, 2015.
 - [17] 満武巨裕. レセプトビッグデータ解析の現状と将来. *実験医学*, Vol. 34, No. 5, pp. 799–804, 2016.
 - [18] 満武巨裕, 大江和彦, 今中雄一. NDB オープンデータを研究利用に活用する：医療技術（CT, MRI, PET）の利用に関する国際比較の試み. *社会保険旬報*, Vol. 2661, pp. 12–16, 2016.