

平成 28 年度 厚生労働行政推進調査事業費研究費補助金（厚生労働科学特別研究事業）
「医療費適正化に向けた生活保護受給者の医薬品処方および生活習慣病の実態調査：大規模レセプト分析」
分担研究報告書

レセプト情報・特定健診等情報データベース

(NDB: National Database of Health Insurance Claims and Specific Health Checkups of Japan)

における仕様と分析対象データの仕様との比較検証

ーデータ利用の活性化を見据えてー

研究分担者 加藤 源太 京都大学医学部附属病院 診療報酬センター 准教授

研究協力者 岩尾 友秀 京都大学医学部附属病院 医療情報企画部 特定研究員

研究要旨

我々が今回利用した医療扶助受給者レセプトデータは、レセプト情報・特定健診等情報データベース（NDB; National Database of Health Insurance Claims and Specific Health Checkups of Japan）とはデータ格納のフローや提供形式が NDB におけるそれとは異なる形式となっている。また、仕様についても NDB をはじめとするレセプトデータの研究に際して参照される記録条件仕様とは異なるものとなっている。本研究では医療扶助受給者レセプトデータの利用経験に基づき、主に「データ提供に至るまでのフロー」「提供ファイルの概要」「データの仕様」について NDB との比較を行うとともに、疫学研究に適したデータ構造への再構成について、我々の取り組みを紹介するものである。

A. 研究目的

今回の調査事業において、我々は厚生労働省社会・援護局を介して生活保護法に基づく医療扶助受給者の全国規模レセプトデータを分析する機会を得た。レセプトデータは、厚生労働省保険局において管理されているレセプト情報・特定健診等情報データベース（NDB; National Database of Health Insurance Claims and Specific Health Checkups of Japan）をはじめとして、保険者や民間が収集したデータベース

が各種研究に活用されており、利用に関する指針やマニュアル、実際に利用した上での成果に関する情報などが徐々に蓄積・公開されており^{1), 2), 3), 4)}、研究者が参照することが可能である。しかし、医療扶助受給者のレセプトデータを全国規模で分析した事例に関する報告はこれまでに見られていない。実際にデータを入手してみると、医療扶助受給者レセプトデータは NDB データが準拠している記録条件仕様とは幾つかの点で異なる構造となっており、構造の解釈や分析のための再構成の作業に時間

を要することとなった。本研究は、医療扶助受給者レセプトデータについて、一定程度データ利用に関する情報が開示されている NDB データと様々な観点から比較を行い、今後このレセプトデータを利用する者にとって参考になる情報を提供することを目的とするものである。

B. 研究方法

【医療扶助受給者レセプトデータと NDB データとの比較事項】

本研究では、医療扶助受給者レセプトデータと NDB データとの間で、「データ提供に至るまでのフロー」、「提供ファイルの概要」および「データの仕様」について比較を行った。

【医療扶助受給者レセプトデータの再構成処理】

上述の NDB との比較を踏まえた上で、医療扶助受給者レセプトデータの特徴を把握し、分析に適した構造に再構成する試みを行った

【倫理面への配慮】

入手した医療扶助受給者レセプトデータは匿名化処理がなされてものではあるが、機微性の高いデータであるという認識のもと、データを扱う PC を個別に準備した。また、それに特定のユーザ名とパスワードを設定し、レセプト解析に携わる研究者のみが使用できるようにした。また、当該 PC はインターネットに接続できない設定を施し、データの流出を防止するデータ分析の

運用体制をとることとした。

C. 研究結果

【データ提供に至るまでのフロー】

NDB では、医療機関から発出されたレセプトデータは、審査支払機関を経由し、保険者から厚生労働省保険局に提出されるが、その際に匿名化処理が行われている。匿名化処理が施された個々のレセプトデータはデータセンターに格納され、利用者からの申出に応じて必要なデータが切り出され、提供される⁵⁾。一方、医療扶助受給者レセプトデータの場合、統計法に基づく一般統計調査として行われている「医療扶助実態調査」にて収集されたデータが対象となっている。平成 25 年より全数データが収集されており、4・5 月診療分、ならびに 6 月審査分の診療報酬明細書及び調剤報酬明細書のうち、一般診療（病院・一般診療所）の入院分及び入院外分、歯科診療分、調剤分の、電子化されたレセプト情報（歯科入院を除く）が収集されている。これらのデータは医療機関から発出された後、社会保険診療報酬支払基金を経由して福祉事務所にて保管されるとともに、福祉事務所において「生活保護等レセプト管理システム匿名化機能」を通じ匿名化され、都道府県・指定都市・中核市においてとりまとめられ厚生労働省社会・援護局に提出される⁶⁾。本研究では、このデータベースに対し目的外利用の申請を行い、厚生労働省社会・援護局からデータの提供を受けた。（図 1）

【提供ファイルの概要】

NDB では、悉皆性が高いデータであることから、一カ月分のサンプリングデータであっても CSV 形式で数 GB 単位の容量となっており⁷⁾、これが年単位、全データとなると数 TB 単位のデータ量となり、きわめて高性能の操作環境と操作技術が利用者には求められる。この点、医療扶助受給者レセプトデータは対象患者数が少ないことと対象時期が限定されていることから、DVD 数枚に収まる程度の容量にとどまっており、利用者の側で極端に性能の高い操作環境を用意する必要はない。それでも、テキストファイルとしては医科全国版で 8.5GB、調剤全国版で 2.3GB となっており、EXCEL など一般的なソフトウェアのなかには開くことができないものがある。限られたソフトウェアでしかファイルが開かない点について、利用者は注意が必要である。(図 2)

【データの仕様】

NDB では、データ構造に関する情報は厚生労働省ホームページ「レセプト情報・特定健診等情報の提供に関するホームページ」に記載されているとともに、その仕様は概ね「オンライン又は光ディスク等による請求に係る記録条件仕様」⁸⁾に準拠しており、利用者はそれら情報を参照しながら研究をすすめることができる。一方、我々が請求提供を受けた医療扶助受給者レセプトデータはそれとは異なる仕様となっていた。各項目のデータ値がいわゆる CSV 形式では一般的なスペースやコンマ等で区切られておらず、複数項目のデータ値が 1 枚のレセプトごとに区切りなしで連続して記載されたものであった。(図 3、図 4、図 5)

以下に、提供された医療扶助受給者レセプトデータのファイルの仕様を示す。

(1) 提供された医療扶助レセプトデータ

【医科全国_確定データ(容量約 9GB)】

下記の項目に関するデータが、スペースやコンマ等の区切りなしで連結され、レセプトごとに 1 行のデータとして格納されている。なお、1 行のデータ数は固定であるものの、下記項目中の傷病名数とサブレコード数はレセプトごとに異なるため、レセプトごとに有効な傷病に関するデータ値数とサブレコード情報の数は異なる。

[医科全国_確定データの項目]

調査年,調査コード,請求年月,診療年月,匿名化 ID 1,匿名化 ID 2,入院・入院外識別フラグ,長期入院識別フラグ,病類 3 区分,カード種別,設置主体識別フラグ,福祉事務所番号,公費負担医療 負担者番号,実施自治体番号,医療機関コード,レセプト種別,後期高齢者識別フラグ,本人識別フラグ,病院・診療所,区分精神,区分結核,区分療養,性別,出生元号,出生年,出生月,年齢,傷病数,主傷病の傷病分類,主傷病の傷病名コード,主傷病の診療開始日,主傷病の ICD-10-1 コード,主傷病の ICD-10-2 コード,主傷病以外の傷病分類(1),診療実日数,入院・診療の元号,入院・診療の年,入院・診療の月,入院・診療期間,本人支払点数,決定点数符号,決定点数,診療行為点数符号,診療行為点数,特定保険医療材料点数,全薬剤総点数,内服薬薬剤総点数,外用薬総点数,注射薬総点数,薬価算出薬剤総点数,薬価算出薬剤内服薬点数,薬価

算出薬剤外用薬点数,薬価算出薬剤注射薬点数,薬価算出薬剤向精神薬点数,後発医薬品の薬価算出薬剤総点数,後発医薬品の薬価算出薬剤内服薬点数,後発医薬品の薬価算出薬剤外用薬点数,後発医薬品の薬価算出薬剤注射薬点数,後発医薬品の薬価算出薬剤向精神薬点数,薬剤種類数,内服薬種類数,外用薬種類数,注射薬種類数,向精神薬種類数,後発医薬品の薬剤種類数,後発医薬品の内服薬種類数,後発医薬品の外用薬種類数,後発医薬品の注射薬種類数,後発医薬品の向精神薬種類数,薬剤処方数,内服薬処方数,外用薬処方数,注射薬処方数,向精神薬処方数,後発医薬品の薬剤処方数,後発医薬品の内服薬処方数,後発医薬品の外用薬処方数,後発医薬品の注射薬処方数,後発医薬品の向精神薬処方数,食事療養費回数,食事療養費金額,生活療養費回数,生活療養費金額,病棟・病院等区分,病院基本料区分,特定入院料区分,初診・再診の区分,再診回数,初診・再診点数,初診・再診回数,医学管理点数,医学管理回数,在宅手技料等点数,在宅手技料等回数,在宅薬剤料点数,在宅薬剤料回数,投薬手技料等点数,投薬手技料等回数,投薬薬剤料点数,投薬薬剤料回数,注射手技料等点数,注射手技料等回数,注射薬剤料点数,注射薬剤料回数,処置手技料等点数,処置手技料等回数,処置薬剤料点数,処置薬剤料回数,手術手技料等点数,手術薬剤料回数,検査手技料等点数,検査手技料等回数,検査薬剤料点数,検査薬剤料回数,画像診断手技料等点数,画像診断手技料等回数,画像診断薬剤料回数,入院点数,入院回数,その他手

技料等点数,その他手技料等回数,その他薬剤料点数,その他薬剤料回数,包括評価部分点数符号,包括評価部分点数,出来高部分点数,包括評価方式,診断群分類番号,除く診療行為の種類数,除く診療行為の点数

【調剤全国_確定データ(容量約 2.5GB)】

同様に、調剤全国_確定データは下記の項目がレセプトごとに 1 行のデータとして格納されたテキストファイルである。

【調剤全国_確定データの項目】

調査年,調査コード,請求年月,診療年月,匿名化 ID 1,匿名化 ID 2,カード種別,設置主体識別フラグ,福祉事務所番号,公費負担医療 負担者番号,実施自治体番号,医療機関コード,識別フラグ,薬局コード,レセプト種別,後期高齢者識別,本人家族,調剤基本科区分,性別,出生元号,出生年,出生月,年齢,受付回数,総点数,調剤行為点数,調剤技術料点数,調剤基本料点数,調剤料点数,薬学管理料点数,特定保険医療材料点数,薬剤料総点数,内服薬薬剤総点数,外用薬総点数,注射薬総点数,薬価算出薬剤総点数,薬価算出薬剤内服薬点数,薬価算出薬剤外用薬点数,薬価算出薬剤注射薬点数,薬価算出薬剤向精神薬点数,後発薬価算出薬剤総点数,後発薬価算出薬剤内服薬点数,後発薬価算出薬剤外用薬点数,後発薬価算出薬剤注射薬点数,後発薬価算出薬剤向精神薬点数,薬剤種類数,内服薬種類数,外用薬種類数,注射薬種類数,向精神薬種類数,後発医薬品の薬剤種類数,

後発医薬品の内服薬種類数,後発医薬品の外用薬種類数,後発医薬品の注射薬種類数,後発医薬品の向精神薬種類数,薬剤処方数,内服薬処方数,外用薬処方数,注射薬処方数,向精神薬処方数,後発医薬品の薬剤処方数,後発医薬品の内服薬処方数,後発医薬品の外用薬処方数,後発医薬品の注射薬処方数,後発医薬品の向精神薬処方数,サブレコード識別フラグ,コード,データ区分+順序番号,点数 x 回数,点数,回数,医薬品の薬価,医薬品の使用量,向精神薬等識別フラグ,医薬品の薬効,医薬品の剤型,処方せん受付回,単位薬剤料,剤型(薬剤),医薬品の後発識別フラグ,薬価算出薬剤点数

(2)提供されたデータの特徴

- 1) 各データ項目の値が、カンマやスペース等の区切りなしで1つのレセプト内で連続した値として格納されている。
- 2) レセプトごとに傷病名や診療行為や薬剤名を表すサブレコード数が可変であり、これらの項目が1行のデータにすべて含まれている。
- 3) 1枚のレセプトごとに1行で記載されているものの、レセプトIDが付与されていない。

(3)提供されたデータの問題点

- 1) データ項目の値に区切りがないため、統計解析ソフトや表計算ソフト等で表形式のデータとして扱うことができず、そのままの形式では解析を行うとができない。これは物理

的なデータ形式の問題である。

- 2) データ解析ではレセプトごとに傷病名やサブレコードを処理する必要がある。解析に適したデータに変換するためには、1行の情報の中に含まれているこれらのデータをレセプトごとに抽出する必要がある。(図6)

【データの再構成処理について】

上記(3)で述べた問題点を解決すること目的に、提供された医療扶助レセプトデータを疫学分析に適した構造に変換するためのソフトウェアを開発することとした。開発にはテキスト整形処理機能が充実しているPython言語を使用することとした。

最終的に、提供ファイルから下記に示すようなデータ解析用の5個のファイルを作成した。

(医科患者背景)

--レセプトIDを付加し、医科の患者背景に関する項目の値を記載したファイル

(医科傷病名)

--レセプトIDを付加し医科傷病名に関するデータ項目値を記載したファイル
(医科サブレコード)

--レセプトIDを付加し、医科サブレコード(診療行為、薬剤名)に関するデータ項目値を記載したファイル

(調剤患者背景)

--レセプトIDを付加し調剤の患者背景に関する項目の値を記載したファイル
(調剤サブレコード)

--レセプトIDを付加し、調剤サブレコ

ード(薬剤名)に関するデータ項目値を記載したファイル(図7、図8、図9、図10)

こうしたデータ再構成処理により、解析対象のデータ項目が表形式のデータとなり、かつレセプトごとにIDが付与されたデータ形式になったため、データの操作性が向上した。(図11、図12)

【その他：患者IDについて】

NDBデータと生活保護レセプトデータとを比較した場合の大きな違いのひとつとして、NDBデータでは氏名の変更や保険者の変更が無い限り原則として患者ごとに一意に定められるIDを備えているが、生活保護レセプトデータでは各受給者の受給者番号が毎月別番号として振られており、受給者番号に由来するIDが月ごとに異なってしまう、という相違があった。これについては、平成29年度よりIDが固定され、月をまたいだ患者単位の分析が可能となった。データ利用者は、平成28年度と29年度とでこのような変更が生じたことに留意する必要がある。

D. 考察

医療扶助受給者レセプトデータには提供時にデータ仕様に関する情報の提供も受けており、それら情報に基づいて丁寧なデータ処理を行えば、データ分析自体は可能である。また、患者IDの精度についてはNDBにおいても改善の必要性が指摘されているが⁹⁾、今後生活保護受給者に対しても各種データを活用して生活習慣病予防等に積極

的に取り組むのであれば、患者単位で一定以上の期間、健康状態や病状を追跡することが必須となるため、より精度の高い患者IDがデータに格納されることが望ましい。

E. 結論

医療扶助受給者レセプトデータについて、NDBデータにおける利用の現状と「データ提供に至るまでのフロー」、「提供ファイルの概要」および「データの仕様」について比較を行った。そのうえで、解析作業に適したデータ構造に再構成を行い、操作性の向上を確認することができた。今後、対象者へのデータヘルスの取り組みを向上させていくためには、利用者の分析用途に適した提供形式としてカンマにより区切られたCSV形式のファイル等での提供、更には信頼度の高い患者IDの格納といった改善が期待される場所である。

G. 研究発表

1. 論文、書籍発表

- 1) 加藤源太. NDB オープンデータを読む際の注意点—いくつかの具体例とともに. 社会保険旬報. 2017:NO.2664:10-5.

2. 学会発表等

- 1) レセプト情報等オンサイトリサーチセンター(京都)の試行的利用に基づく今後の活用可能性に関する検証(シンポジウム)、酒井未知、大寺祥祐、岩尾友

- 秀、岡本和也、加藤源太、中山健夫、黒田知宏、第 36 回医療情報学連合大会、2016/11/23, 横浜.
- 2) NDB(ナショナルデータベース)の課題および留意点と今後の展望、久保 慎一郎、野田 龍也、明神 大也、加藤 源太、今村 知明、第 36 回医療情報学連合大会、2016/11/24, 国内.
- 3) HOMAS2 を利用した後発医薬品切替への取り組み—ベンチマークデータの活用—、駒井和子、寺前利治、首藤真義、服部敦史、谷口陽平、數藤大輔、中村文子、佐藤里香子、松原和夫、深津祥央、尾崎淳子、田村寛、加藤源太、黒田知宏、平家俊男、平成 28 年度大学病院情報マネジメント部門連絡会議、2017/01/27, 国内.
- 4) データを活用した病院経営への取組事例—病院運営企画室の活動紹介—、服部敦史、寺前 利治、首藤 真義、駒井 和子、小谷 将司、杉野 剛史、加藤 源太、田村 寛、黒田 知宏、平家 俊男、宮本 享、平成 28 年度大学病院情報マネジメント部門連絡会議、2017/01/27, 国内.
- 5) レセプトデータの二次利用について：マクロ (NDB)・ミクロ (院内データ) 双方の視点から、加藤源太、関西医療情報処理懇談会 第 54 回例会、2017/01/28, 国内
- 6) 海外におけるレセプトデータ利活用の支援について—米国 ResDAC の取り組みから—、加藤源太、日本疫学会 第 14 回医療情報・統計利用促進委員会例会、2017/03/15, 国内

H. 知的財産権の出願・登録状況 (予定を含

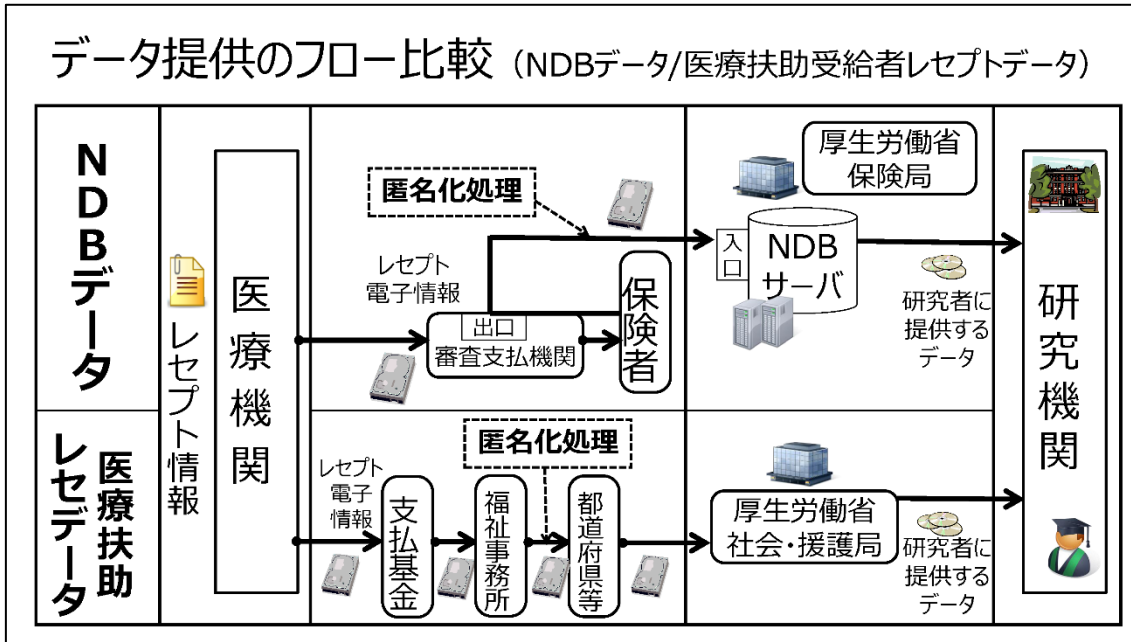
む)

- | | |
|-----------|----|
| 1. 特許取得 | なし |
| 2. 実用新案登録 | なし |
| 3. その他 | なし |

引用・参考文献

- 1) 厚生労働省、「レセプト情報・特定健診等情報の提供に関するホームページ」、http://www.mhlw.go.jp/stf/seisakunit suite/bunya/kenkou_iryuu/iryuuho ken/reseputo/index.html、(2017.03.31 確認)
- 2) 厚生労働省、「レセプト情報・特定健診等情報データベースの第三者提供 — 利用を検討している方々へのマニュアル」、<http://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000117728.pdf>、(2017.03.31 確認)
- 3) 厚生労働省、「第 36 回レセプト情報等の提供に関する有識者会議 第三者提供の成果物集計について」、<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000155475.pdf>、(2017.03.31 確認)
- 4) 加藤源太, 平野景子, 赤羽根直樹. レセプト情報・特定健診等情報データベースの利活用について. 統計, 2014;65(10):8-13.
- 5) 厚生労働省、「第 1 回レセプト情報等の提供に関する有識者会議 レセプト・特定健診等データについて」、<http://www.mhlw.go.jp/stf/shingi/2r9>

- 852000000thao-
att/2r985200000theb.pdf、
(2017.03.31 確認)
- 6) 厚生労働省、「医療扶助実態調査：調査の概要」、
<http://www.mhlw.go.jp/toukei/list/67-16a.html#link01>、
(2017.03.31 確認)
- 7) 厚生労働省、「第9回レセプト情報等の提供に関する有識者会議 基本データセットについて」、
<http://www.mhlw.go.jp/stf/shingi/2r98520000024d12-att/2r98520000024d5x.pdf>、
(2017.03.31 確認)
- 8) 診療報酬情報提供サービス、「レセ電システムに関する情報」、
<http://www.iryohoken.go.jp/shinryohoshu/receMenu/doReceInfo>、
(2017.03.31 確認)
- 9) 厚生労働省、「第28回レセプト情報等の提供に関する有識者会議 レセプト情報と特定健診等情報の突合について」、
<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000115075.pdf>、
(2017.03.31 確認)



【図1：データの格納・提供に係るフロー】

提供ファイルの概要

- 医科全国_確定データ (File size: 8.5GB)
- 調剤全国_確定データ (File size: 2.3GB)

} × EXCEL

} × サクラエディタ

○ Em Editor

テキストファイルのサイズが大きいため、限られたソフトウェアでしか開くことができない

【図2：提供ファイルの概要及びソフトウェアでの対応】

医療扶助受給者レセプトデータ ファイル構造(例：医科全国データ)



【図3：医療扶助受給者レセプトデータのファイル構造】

参考：ファイル構造(NDB)

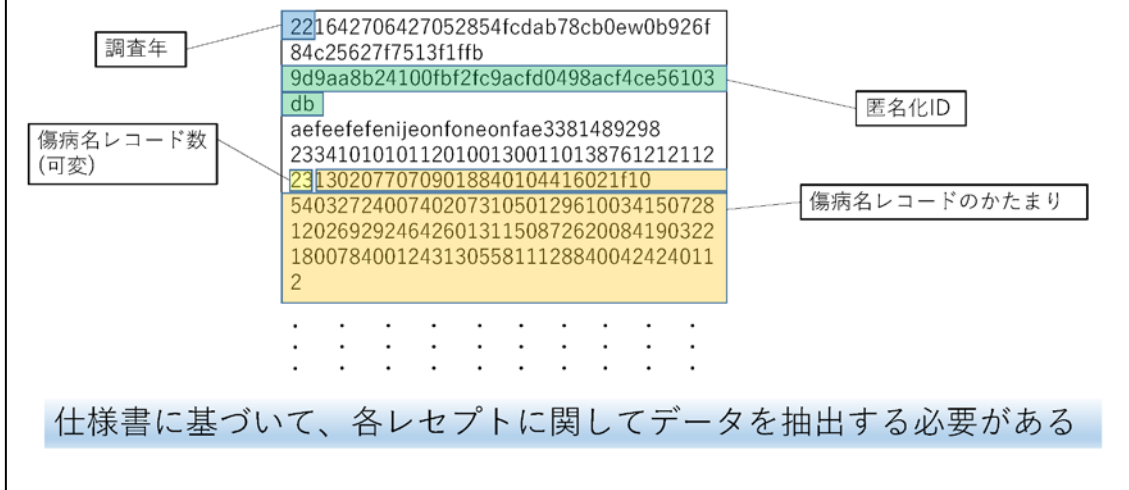
レコード順序：レセプト行の絶対順位（この番号を用いてソートが可能）
レセプト通番：同一レセプトには同一番号が振られており統合が可能
レコード識別名：どの種類の情報かを示す（例：SY→傷病名）

患者ID

5108	RECODEINFO_MED_SAMPLE_6000000174	IR, 5.1.1234567.42303, , , , ,
5109	RECODEINFO_MED_SAMPLE_6000000174	RE 37c2897fa253
5110	RECODEINFO_MED_SAMPLE_6000000174	HC, 11111111, 31, 70759, 0, 0, 44400, , , ,
5111	RECODEINFO_MED_SAMPLE_6000000174	SY 8839792, 4061031, 1, , , , , , , ,
5112	RECODEINFO_MED_SAMPLE_6000000174	SY 1519006, 4221119, 3, 3088, , , , , , , ,
5113	RECODEINFO_MED_SAMPLE_6000000174	SY 999, 4220902, 1, , , , , , , ,
5114	RECODEINFO_MED_SAMPLE_6000000174	SY 9970003, 4221215, 3, , , , , , , ,
5115	RECODEINFO_MED_SAMPLE_6000000174	SY 3543005, 4221015, 1, 2056, , , , , , , ,
5330	RECODEINFO_MED_SAMPLE_6000000177	IR, 2.10.1.1234567.42303, , , , ,
5331	RECODEINFO_MED_SAMPLE_6000000177	RE 31f5c9f40433f
5332	RECODEINFO_MED_SAMPLE_6000000177	HC, 11111111, 4, 8434, 4, 2560, 8410, , , ,
...		

【図4：NDBデータのファイル構造】

1行分のデータ形式(医科全国データ)



【図5：1枚分のレセプトデータに相当する、1行分のデータ形式の例】

ファイル構造の問題点

[レセプト解析における問題点]

- 1行分のデータに、可変長のデータ(傷病関連等)が含まれている
- レセプト番号が付与されていない

[計算機処理における問題点]

- データ項目間が「カンマ」や「スペース」で区切られていない
(現実：4250415727583758⇒理想：4250415, 72, 7583,75,8)
⇒Excelや統計解析用ソフトウェアでは扱うことができない
(≡ データが持っている意味を解釈をすることができない)

【図6：医療扶助受給者レセプトデータのファイル構造における問題点】

提供ファイルの改修方針

[レセプト解析における問題点の解決]

- 1行分のデータに含まれている可変長のデータ(傷病関連等)を、別ファイルとして作成する
- 1行分のデータごとにレセプト番号を付与する

[計算機処理における問題点の解決]

- 仕様書に基づいて、データ項目間に「カンマ」を挿入する

改修方法：

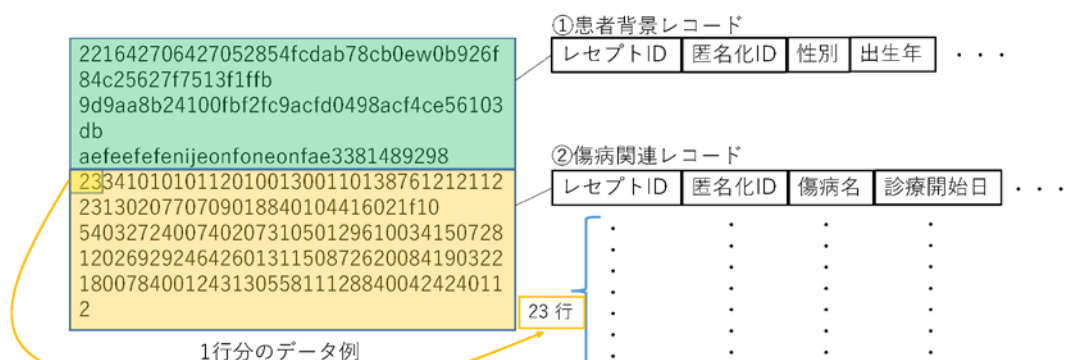
手作業⇒×(∵データ量が多いため、現実的ではない)

Excel VBA⇒×(∵処理速度が遅く、かつ複雑な処理には不向き)

プログラミング言語⇒◎(今回は、Pythonを使用)

【図7：提供ファイルの改修方針】

データ再構成の実施(例：医科全国データ)



①「患者背景」と②「傷病関連レコード群」を別ファイルに分割

【図8：データ再構成の実施イメージ】

データ再構成により作成したファイル

作成 ファイル名	医科 患者背景	医科 傷病名	医科 サブ レコード	調剤 患者背景	調剤 サブレコード
レコード構造	縦型	縦型	縦型	縦型	縦型
提供 ファイル名	医科全国_確定データ			調剤全国_確定データの項目	

【図9：データ再構成により作成したファイル】

1行分のデータ形式(医科全国データ)

調査年

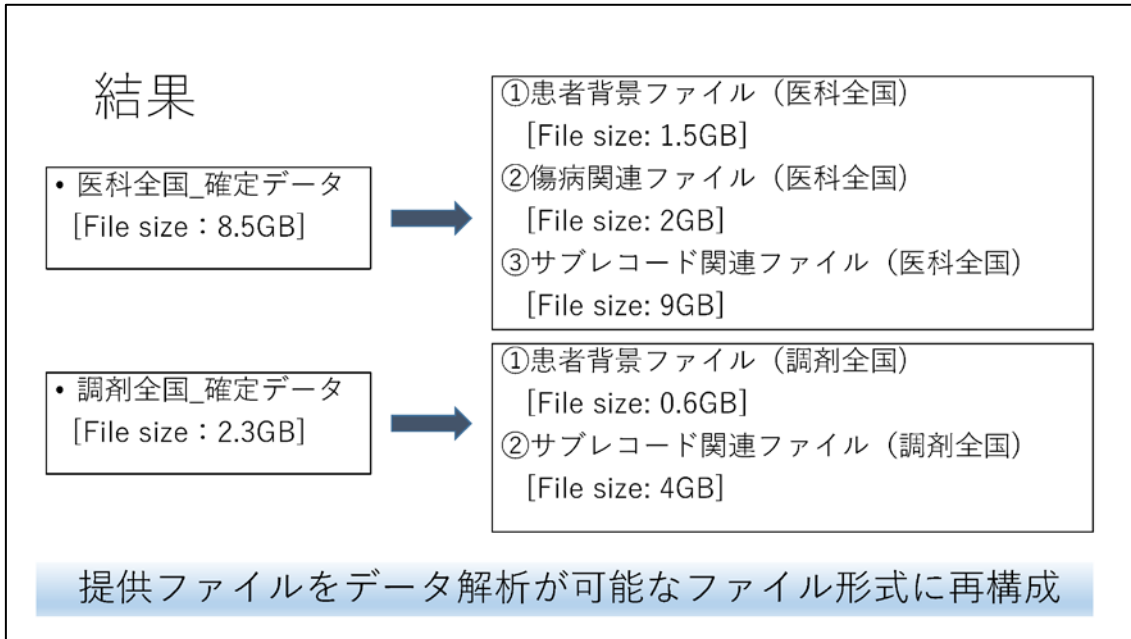
匿名化ID

サブレコード数(可変)

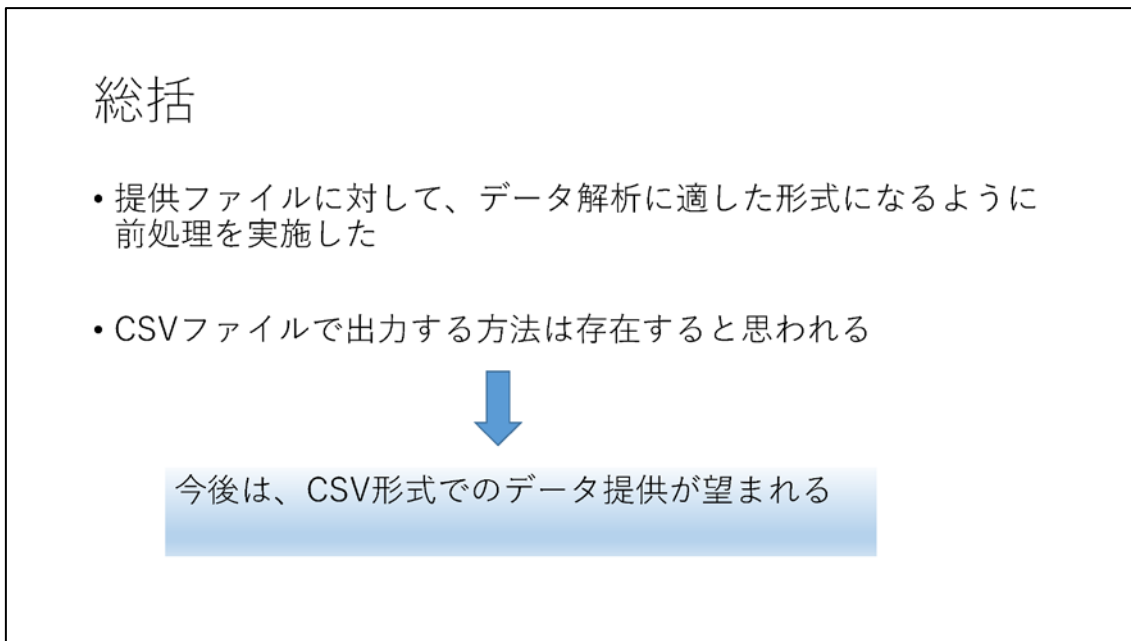
サブレコードのかたまり(診療行為、医薬品等)

1行分(1レセプト)データには、サブレコードも含まれている

【図10：データ形式】



【図 11 : データ再構成化の結果】



【図 12 : 総括・CSV化の必要性】