

ビッグデータ分析のための レセプト(診療報酬明細書)データ活用例の報告

Education of analytic skills of healthcare-related “big data” using health insurance claims data

岡本悦司, 神谷達夫, 北口千華, 大槻知宏

Etsuji Okamoto, Tatsuo Kamitani, Chihana Kitaguchi, Chihiro Otsuki

要旨

ビッグデータ活用の重要性が高まっており、ビッグデータを分析する技能(スキル)の習得が、文系も含む全ての大学教育に求められている。とりわけ医療分野は、レセプト電子化や病院情報の公開もあいまって、医療ビッグデータの積極的な活用が国家戦略としても位置づけられるようになった。同時にまた、医療データは単にデータサイズが大きだけでなく、きわめて複雑かつ専門的な内容を含むことが教育上の試練となる。このたび教育用に購入された匿名加工レセプトデータを用いて医療福祉経営学科の専門教育への活用を試みたので報告する。

キーワード: 医療ビッグデータ, レセプト(診療報酬明細書), SQL, 日本医療データセンター社 (JMDC)

Keywords: healthcare-related big data, health insurance claims, SQL, Japan Medical Data Center

1. はじめに

政府が示す AI 戦略 2019[1]では「すべての大学・高専生(約 50 万人/年)が初級レベルの数理・データサイエンス・AI を習得すること」を目標に掲げており、IT スキルは、今や情報系のみならず文系を含む全ての大学教育に求められる技能となった。文部科学省も「数理・データサイエンス・AI 教育プログラム」認定制度を 2020 年度よりスタートさせ、福知山公立大学地域経営学部もリテラシーレベルの認定を受けている。

しかしながら、医療データは、その構造が複雑かつ専門的な内容を含むので、リテラシーレベルの科目ではその分析技能を習得することはできない。医療福祉経営学科で

は、4年生対象の専門科目として「医療事務総論」を開講しているが、この科目は医療機関が作成するレセプト(診療報酬明細書)の構造とデータ分析法を習得する科目である。教育法は講義と PC 演習を併用し、サンプルデータを用いてレセプトデータの構造を理解させることで実施してきたが、数理・データサイエンス・AIセンターが教育用に匿名加工が施されたレセプトデータを日本医療データセンター(JMDC)社より購入したので、そのデータ(以下、JMDC データ)を用い、科目受講学生と共に、専用サーバーに取り込み、学生が主体となったデータ分析に取り組めるように加工した。

2. レセプトデータの構造

レセプトは正式には診療報酬明細書と呼ばれ、保険が適用される、いわゆる保険診療を提供した後、患者から徴収する一部負担金をした残りを医療機関が保険者に請求するために送付する請求書の一部である(明細書は患者単位に作成され、請求書はその暦月に医療機関が請求する総額である)。

レセプトは暦月単位に作成・提出され、1件、2件と数えられる。レセプトは、医療機関の種類によって医科、歯科そして調剤に大別され、医科はさらに入院と入院外(外来と在宅)そして一部の急性期病院が対象となる DPC(diagnosis-procedure-combination, 診断群分類による定額払い)の3種に区別される。

レセプトはごく最近まで紙で提出されていたが、最近では大半が電子化されている。オンラインで提出されるレセプトデータは、csv形式のテキストファイルであり、基本的には、紙レセプトの記載内容をそのままテキストデータに変換したかたちになっている【図1】。

【図1】紙レセプトと電子レセプト(csv形式)の対応

電子レセプト (レセコードデータ)

IR,1.13.1.9999996.基金病院.41802.00...
 RE,55.1116.41711.基金 花子.2.3220725.....199.1116.55...01
 H,0.06139992.55.55.1.144 H0は保険者コードFE記号・番号

SY,8838506.4120824.1... SYは病名名、7桁の病名コード診療
 SY,8840829.4120824.1... 要約日が記載ある、末尾の0は主病名
 SY,2724007.4121116.1... 0002は修飾語(無い)、2036G(石)を示す

SY,5319009.4131117.1...
 SY,8835582.4131212.1.2056..
 SY,2500013.4140131.1...01
 SY,2500014.4171005.1.8002..

SI,12.1.112007410.58.1
 SI,12.1.112011010.52.1
 SI,12.1.112007170.5.1
 SI,13.1.113002010.87.1
 SI,13.1.120002370.10.1
 SI,21.1.120000710.9.1
 IY,21.1.61043013.2.000.16.56
 SI,25.1.120001210.42.1
 SI,25.1.120003170.45.1
 SI,27.1.120001810.8.1
 SI,60.1.160095710.12.1
 SI,60.1.160061910.155.1
 SI,60.1.160022410...
 SI,1.160020910...
 SI,1.160023410...
 SI,1.160020810.62.1
 G0,1.1441.99 G0は合計のコード

○厚生労働大臣の定める規格及び方式
 ○レセプト電算処理マスターコードで記載
 病名名マスター
 診療行為マスター
 医業記マスター
 特定药材マスター
 コメントマスター

2.1 レコードとレコード識別

レセプトは、診療録に基づいて作成され、傷病名、提供された医療行為そして医薬品や特定機材等の全ての情報を含む。電子レセプトでは各行を「レコード」と呼び、各レコードが、それぞれ異なるデータを含んでいる。各行の左端に「レコード識別」と呼ばれるアルファベット 2 文字が含まれており、レコード識別によってそのレコードがどの情報を含むか判別する。

電子レセプトに含まれるレコードには以下のようなものがある。

2.1.1 レセプトの基本情報

これはレセプトごとにそのレセプトの基本情報を含むレコードであり以下の 3 つがある。これらは基本的に各レセプトに 1 レコードのみである。

- ・医療機関レコード(IR:医療機関のイリ)・・・医療機関コード(10 ケタ)や医療機関名称等
- ・レセプト共通レコード(RE:レセプトのレ)・・・患者氏名、生年月日、性、診療年月
- ・保険者レコード(HO:ホケンシャのホ)・・・保険者コード(8 ケタ)、請求点数や診療実日数

2.1.2 傷病名情報

レコード識別は SY(ショウビョウのショ)であり、傷病名コード(7 ケタ)の他、診療開始日、転帰、主傷病フラグ等が含まれる。記載する傷病名の数に制限はないので高額レセプトでは時には 100 を超えるレコードになることもある。

2.1.3 摘要欄情報

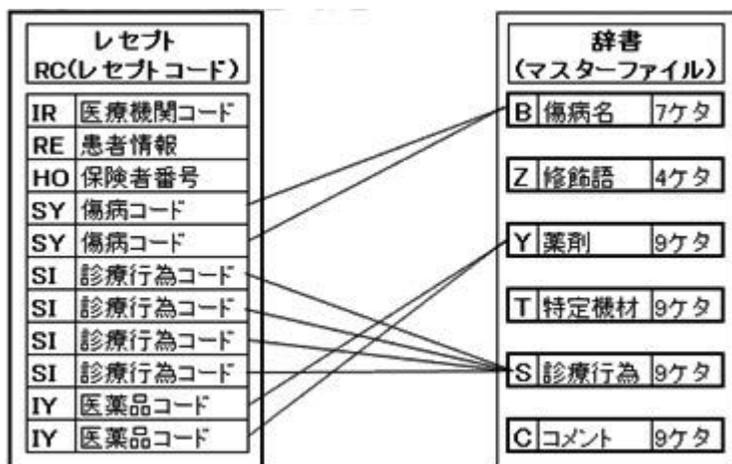
摘要欄とは、投薬ならどの医薬品をどれだけ投与したか、診療行為なら、どのような治療を何回提供したか、という医療の詳細な内容を含む、いわばレセプトの心臓部にあたる。これには以下のようなレコードが含まれる。

- ・診療行為レコード(SI: シンリョウのシ)・・・診療行為コード(9 ケタ)と数量、点数、回数等
- ・医薬品レコード(IY:イヤクヒンのイヤ)・・・薬剤コード(9 ケタ)と数量、点数、回数等
- ・特定器材レコード(TO:トクテイのト)・・・特定器材コード(9 ケタ)と数量、点数、回数等

2.2 各種マスター

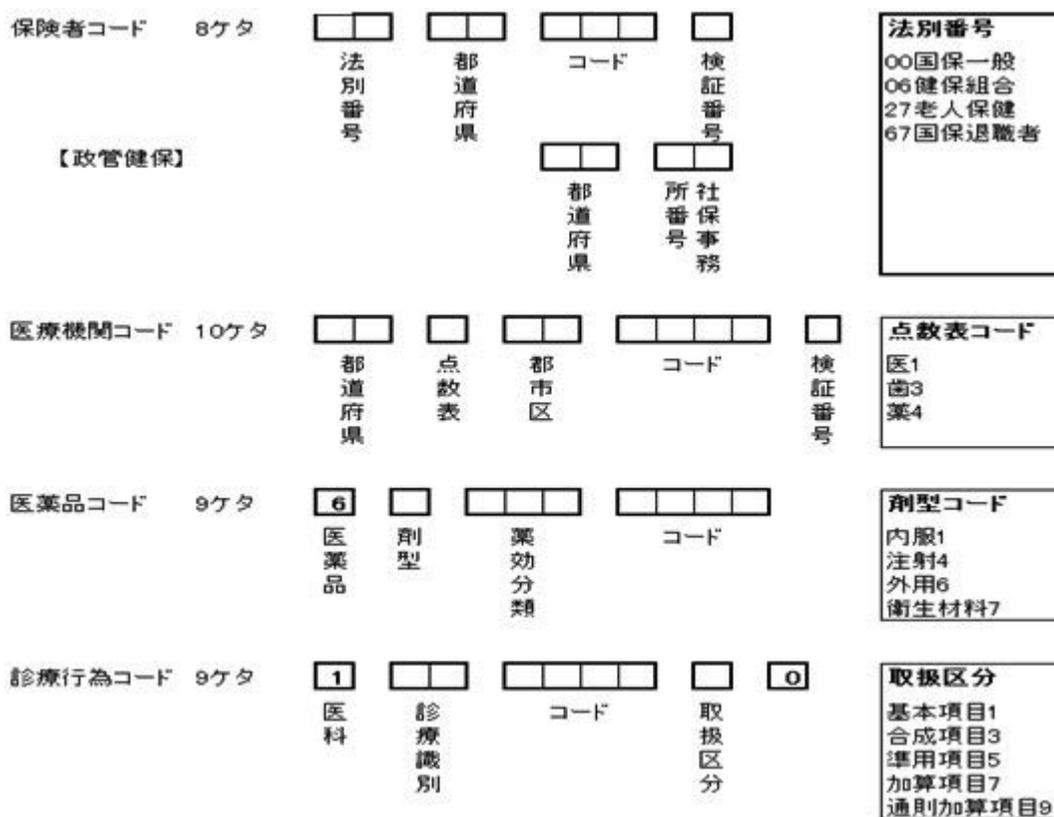
レセプトに含まれるデータは原則として数字コードで表記される。よって、レセプトデータ分析を行うには、必ず、数字コードが何を意味するかを参照する「辞書」すなわち「マスター」とリンクする必要がある。たとえば、医療機関コードで 2619600212 は福知山市民病院を、保険者コードで 01260017 は全国健康保険協会京都府支部を、それぞれ意味する[2]。

【図2】電子レセプトの構造と辞書(マスター)との関連



レセプトは多様なコードが含まれるが、それらコードには一定の構造があり、それを理解することは効率的なレセプト分析に役立つ。

【図3】レセプトコードの構造



3. JMDC データの概要

日本医療データセンター(JMDC)社は、医療データを専門とする民間企業であり、主として健康保険組合と契約してレセプトデータの提供を受け、匿名加工を施して製薬企業等にデータを販売するビジネスモデルをとっている。その際、単に個人情報を匿名化するだけでなく、データのクリーニングや正規化等の加工を行い、データを分析しやすいかたちにして提供している。

数理・データサイエンス・AI センターが購入したデータ(以下、JMDC データ)はファイルサイズが全体で 31GB、最大の診療行為ファイルが 18.3GB、医薬品ファイルは 7.4GB、傷病ファイルは 4.4GB であった。

これまでデータウェアハウス化においては、ACCESS に読み取り、それを Excel の PowerPivot 機能を用いてピボットテーブル操作できるよう加工することで行ってきた[3]。そのメリットは第一にファイルサイズの圧縮である。ACCESS に取り込んで正規化(関係データベースにおいてデータの冗長性をなくす処理)を行うことによって、ファイルサイズを約 10 分の 1 に圧縮することができる(たとえば 1GB の ACCESS ファイルを Excel の PowerPivot に読み込むと Excel のファイルサイズは約 100MB くらいですむようになる)。それでも ACCESS は 2GB が限界であるため、JMDC データを取り込むことはできず、処理には、専用サーバーを構築し、MySQL によって操作することとした。

JMDC データは約 1750 万件のレセプト、ユニークな患者数約 349 万人を含むビッグデータである。含まれるのは入院、入院外(外来と在宅)そして DPC の 3 種類の医科レセプトのみであり、調剤や歯科レセプトは含まれてはいない。また、生年月や性別といった患者情報は全体の 18.8%程度しかない点が限界ではある。さらに、データの出所等は明らかにされておらず、あくまで教育用データであって、その内容に関する保証はないことも留意する必要がある。

【表 1】 JMDC データの概要

	レセプト件数[R]			患者数(ユニーク)[P]		
	うち患者情報有[r]	r/R		うち患者情報有[p]	p/P	
DPC	1349190	253447	18.8%	712711	133654	18.8%
入院	592161	111523	18.8%	279786	52708	18.8%
入院外	15559962	2920111	18.8%	3261735	611671	18.8%
全体	17501313	3285081	18.8%	3490596	654499	18.8%

3.1 データの構造

レセプトデータは、前述のように、様々なレコードがレセプトごとに繋がっている。そのため、分析のためには、同フォーマットのレコード識別ごとに区分し、どのレコードがどのレセプトのレコードであるかを識別する識別子(レセプト ID)を付加する必要がある。レセプト ID があれば、同一レセプトに記載された傷病レコードと医薬品レコードを結合して、どの傷病に対してはどのような医薬品が投与されたか、が分析可能となる。

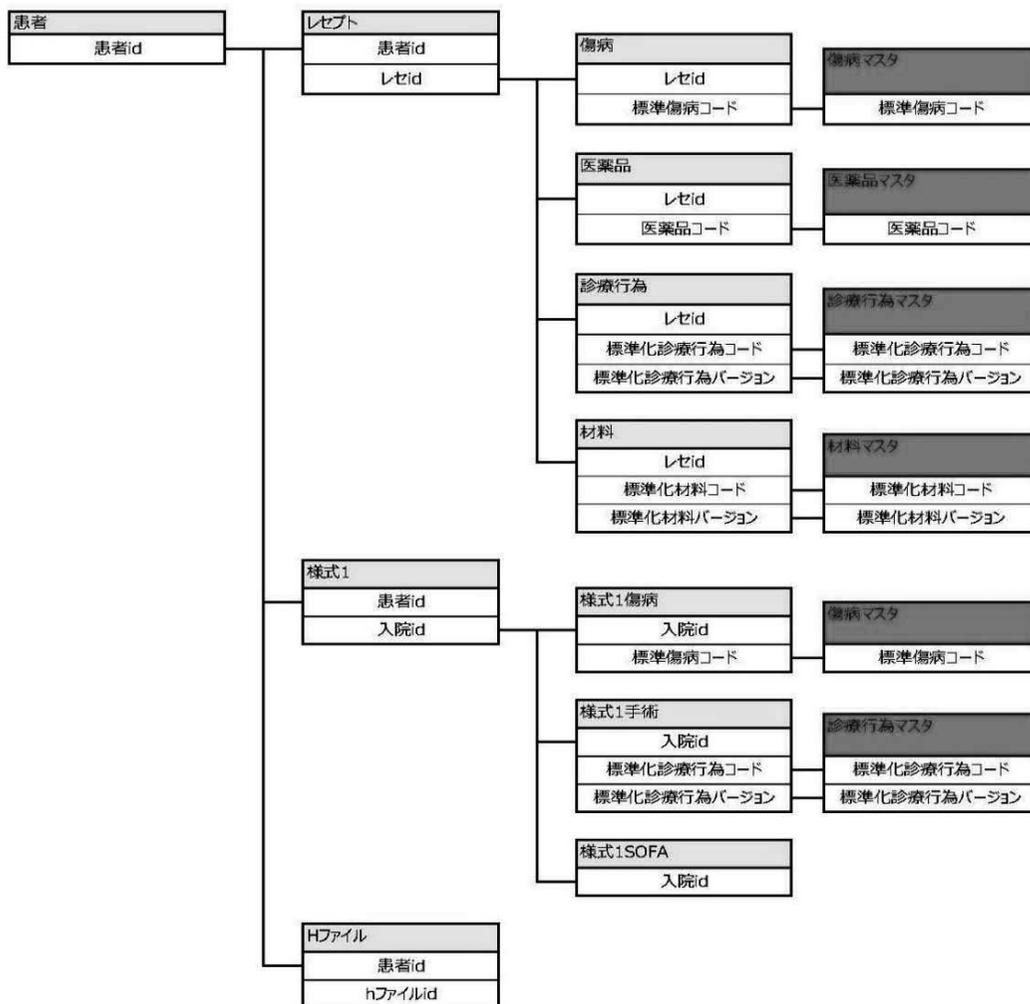
さらに、医薬品は薬効で分類され、薬効には大分類、中分類、小分類といった段階がある。大分類→小分類へと分類を細かくすることをドリルダウン、その逆をドリルアップと呼びデータウェアハウスの重要な機能である。こうした薬効分類を各医薬品レコードに含めるのではなく、医薬品レコードには医薬品コードのみとして、別に医薬品マスターファイルを作り、医薬品コードと大・中・小分類情報を結合させれば冗長性を減らす(=ファイルサイズを小さくし、検索を早くする)ことができる。

レセプトはまた、医療機関ごとに、種類(入院、入院外等)ごとに暦月単位で作成されるので、同一の個人から複数のレセプトがでてくる。そこでどのレセプトが誰に属するのか、の関係を示す患者 ID とレセプト ID を関連づける対応表も必要となる。患者 ID と結合されることによって、性別や年齢別といった患者属性によって傷病の出現率や医薬品処方状況の違いを出すことができる。

以上のような、雑多なデータを種類ごとのファイルに分け、異なるファイル間を ID 連結するよう加工する作業を正規化と呼ぶが、正規化作業はレセプトデータ分析で最初に行わなければならない処理であり、かつ最も手間のかかる作業である。JMDC データは予め、この正規化処理が行なわれた後で提供される。すなわち最初から、レセプトと患者との対応表、診療行為や医薬品ファイルに加工されており、さらに診療行為や医薬品のマスターファイルも提供されている。その概要は【図 4】の通りである。

図中「レセプト」ファイルはレセプト ID と患者 ID との対応表であり、「様式 1」とは DPC(診断群分類)適用の入院患者についての退院時抄録(discharge summary), H ファイルとは DPC 適用の入院患者についての「看護必要度」データである。また「様式 1」とつながる SOFA(sequential organ failure assessment score)とは患者の重症度を指す。

【図 4】JMDC データの正規化の結果



3.2 診療月別、レセプト種類別の概要

診療月別にみると 2020 年 1 月～2021 年 12 月の 24 か月間であり、概ね新型コロナ流行の時期と重なる。そのため、件数も月別に変動がみられ、2020 年 5 月や 2021 年 2 月に件数の減少がみられる。また患者情報有りの割合も月によって若干の差異がみられる。

【表 2】 JMDC データの診療月別件数と患者情報あり割合

診療年月	レセプト件数				患者情報のある割合			
	DPC	入院	入院外	計	DPC	入院	入院外	計
202001	58327	25768	703999	788094	18.7%	18.8%	18.7%	18.7%
202002	59430	26435	645538	731403	18.7%	18.9%	18.8%	18.8%
202003	59055	26389	667187	752631	18.8%	19.1%	18.8%	18.8%
202004	51525	23846	585928	661299	18.7%	19.2%	18.7%	18.7%
202005	47129	22329	536931	606389	18.8%	18.6%	18.7%	18.7%
202006	54468	23399	630923	708790	18.7%	18.6%	18.8%	18.8%
202007	58265	24362	657441	740068	18.7%	18.7%	18.8%	18.8%
202008	58612	24698	636254	719564	18.8%	18.8%	18.7%	18.7%
202009	56842	24390	652947	734179	18.7%	19.1%	18.8%	18.8%
202010	60919	25300	686117	772336	18.8%	18.7%	18.7%	18.7%
202011	57718	24976	637136	719830	18.9%	19.0%	18.8%	18.8%
202012	57537	25855	667568	750960	18.7%	19.1%	18.7%	18.7%
202101	52626	24794	617508	694928	18.6%	18.4%	18.8%	18.8%
202102	52641	24253	593658	670552	18.9%	18.7%	18.8%	18.8%
202103	57686	25258	703793	786737	18.9%	18.7%	18.8%	18.8%
202104	56823	25339	660241	742403	18.7%	18.7%	18.8%	18.8%
202105	53411	24916	603502	681829	18.7%	18.9%	18.8%	18.8%
202106	56611	24745	666401	747757	18.9%	19.0%	18.8%	18.8%
202107	57141	24933	664618	746692	18.8%	18.9%	18.8%	18.8%
202108	57279	26449	669540	753268	18.6%	19.0%	18.7%	18.7%
202109	54535	24725	664970	744230	19.0%	18.8%	18.8%	18.8%
202110	58037	23748	671329	753114	19.0%	18.7%	18.8%	18.8%
202111	57603	23072	656982	737657	18.8%	18.8%	18.7%	18.8%
202112	54970	22182	679451	756603	18.7%	19.0%	18.8%	18.8%
計	1349190	592161	15559962	17501313	18.8%	18.8%	18.8%	18.8%

3.3 マスターファイル

傷病名、診療行為そして医薬品といった各種コードの「辞書」にあたるマスターファイルは厚生労働省サイトより提供されており、【図 3】のように各種コードは決められた桁数からなっている。

JMDC データは厚生労働省が公開するマスターファイルの一部を修正した独自のマスターファイルを含んでいるので、JMDC マスターファイルに含まれる項目とそのフォーマット等を、具体例も含めて説明する。

3.3.1 傷病マスター

レセプトデータに記入される傷病コードは7ケタで、その上4ケタはICD9コーディングである[4]。たとえば「0050002」の上4ケタである005.0は「ブドウ球菌性食中毒」を指す。上3ケタの005は「細菌性食中毒」を指す。対して「ブドウ球菌感染症」は041.1なので「0411001」となる。ICD9は1994年まで使われていた古い分類であるが、電子レセプトのフォーマットが最初に定められたのは1991年だったのでICD9が使われ、傷病コードは今も変わっていない。ICDは1995年以降現在に到るまでICD10が使われるようになったので、傷病マスターはいわばICD9→ICD10の対応表のようにになっている[5]。

【表3】傷病マスター

傷病マスタ				
通番	項目名	フィールド名	フォーマット	具体例
1	標準傷病コード	sycode	char(7)	0050002
2	標準病名	syname	varchar(256)	ぶどう球菌食中毒
3	ICD10大分類コード	icd10daicode	char(7)	A00-B99
4	ICD10大分類名	icd10dainame	varchar(256)	感染症及び寄生虫症
5	ICD10中分類コード	icd10chucode	char(7)	A00-A09
6	ICD10中分類名	icd10chuname	varchar(256)	腸管感染症
7	ICD10小分類コード	icd10shocode	char(3)	A05
8	ICD10小分類名	icd10shoname	varchar(256)	その他の細菌性食中毒, 他に分類されないもの
9	ICD10細分類コード	icd10csaicode	char(6)	A050
10	ICD10細分類名	icd10sainame	varchar(256)	ブドウ球菌性食中毒

3.3.2 診療行為マスター

診療行為コードは医薬品とも同じ9ケタの数字である。最初の1ケタは医科では「1」である(歯科は「3」、調剤は「4」そして医薬品は「6」)[6]。

【表4】診療行為マスター

診療行為マスタ					
通番	項目名	説明	フィールド名	フォーマット	具体例
1	診療行為コード	診療行為コード	sicode	char(9)	111011570
2	診療行為バージョン	診療報酬改定ごとのバージョン	siver	char(6)	202004
3	診療行為名	診療行為名称	meishou	varchar(256)	乳幼児夜間加算(小児科初診)
4	診療区分中分類名	診療点数早見表の目次中分類	sidai	varchar(256)	初・再診料
5	診療区分小分類名	診療点数早見表の目次小分類	sichu	varchar(256)	初・再診料
6	診療区分細分類名	診療点数早見表の目次細分類	sisho	varchar(256)	初診料
7	診療区分細々分類名	診療点数早見表の目次細々分類	sisai	varchar(256)	
8	診療点数早見表区分コード	診療点数早見表のアルファベット及び数字	sikubuncode	char(8)	A000
9	icd9cm_level1	ICD-9CM大分類	icd9cm1	char(5)	
10	icd9cm_level2	ICD-9CM中分類	icd9cm2	char(5)	
11	icd9cm_level3	ICD-9CM小分類	icd9cm3	char(5)	

3.3.2.1 診療識別

2～3ケタ目は「診療識別」と呼ばれるもので、診療の種類を表す【表5】。たとえば「111011570」は医科の初診(11)に属する診療行為であることがわかる。

【表5】診療識別(診療行為コードの2～3桁目)

11	初診	26	麻酔	50	手術・麻酔
12	再診	27	調基	54	麻酔
13	医学管理	28	その他	60	検査
14	在宅	30	注射	70	画像診断
20	投薬	31	皮下筋肉	80	その他
21	内服	32	静脈	90	入院
22	頓服	33	その他	92	特定入院料
23	外用	39	薬剤料減点	93	診断群分類
24	調剤	40	処置	97	食事・生活
25	処方			99	全体に関わる識別コード

3.3.2.2 取扱い区分

診療行為コードの8ケタ目は「取扱い区分」と呼ばれ、診療行為の種類を示す。基本項目とは単独で請求される診療行為であり、加算項目とは、たとえば乳幼児が小児科を夜間に受診した場合に、通常の前診料に加算して請求される点数を指す。具体例では8ケタ目が「7」であることより(単独では請求されない)加算項目であることがわかる【表6】。

【表 6】取扱い区分

診療行為コード8桁目	取扱い区分	診療行為コードの数
1	基本項目	2076
3	合成項目	39
5	準用項目	420
7	加算項目	332
9	通則加算項目	65
		2932

診療行為マスターは、ドリルダウン、ドリルアップが可能なように、中分類、小分類、細分類が記載されている。大分類がないのは、区分コードのアルファベット部分が大分類を示すためである。たとえば 111011570 は大分類「基本診療料」、中・小分類「初・再診料」そして細分類は「初診料」に属するものであることがわかる【表 7】。[7]

【表 7】診療行為大分類

診療行為大分類	
A 基本診療料	H リハビリテーション
B 医学管理	I 精神科専門療法
C 在宅医療	J 処置
D 検査	K 手術
E 画像診断	L 麻酔
F 投薬	M 放射線治療
G 注射	N 病理診断

JMDC の診療行為マスターには ICD9CM 分類も付加されている。ICD9 とは国際疾病分類のことであるが、アメリカでは ICD9-CM (clinical modification) という手術や処置といった診療行為に関する分類も制定しており、JMDC は手術及び処置といった診療行為大分類については ICD9-CM も付記している[8]。

3.3.3 医薬品マスター

厚生労働省が提供する医薬品マスターにはレセプトに記載されるレセプト電算コード(9ケタ)に加えて、薬価基準(YJ)コード(12ケタ)も含まれており、薬効分類目的では通常こちらを使用する。JMDC は主に製薬企業をユーザーとすることもあり、その医薬品マスターは詳細で、国際的に用いられる ATC 分類も付記されている。

【表 8】医薬品マスター

【表】医薬品マスタの構成

項目名	説明	フィールド名	フォーマット	具体例
1 医薬品コード	医薬品に振られている、JMDC独自のユニークID	phma_prod_cd	char(24)	100000000378
2 医薬品名	各製薬会社の販売名	drug_name	varchar(1200)	コナン錠 1 0 m g
3 規格単位	薬価算定のための単位	std_unit	varchar(800)	1 0 m g 1 錠
4 ATC大分類コード	解剖学的部位に基づいた分類[アルファベット1文字]	atc_lv1_cd	char(2)	C
5 ATC大分類名		atc_lv1_name	varchar(400)	循環器系用薬
6 ATC中分類コード	治療法メイングループによる分類[2個の数字]	atc_lv2_cd	char(6)	C09
7 ATC中分類名		atc_lv2_name	varchar(800)	レニン・アンジオテンシン系作用薬
8 ATC小分類コード	治療法・薬学サブグループによる分類[1個のアルファベット]	atc_lv3_cd	char(8)	C09A
9 ATC小分類名		atc_lv3_name	varchar(800)	A C E 阻 害 薬 単 味 剤
10 ATC細分類コード	化学・治療法・薬学サブグループによる分類[1個のアルファベット]	atc_lv4_cd	char(10)	C09A-
11 ATC細分類名		atc_lv4_name	varchar(800)	A C E 阻 害 薬 単 味 剤
12 ATCコード	化学構造サブグループによる分類[2個の数字]	who_atc_cd	char(14)	C09AA06
13 ATC名		who_atc_name	varchar(1600)	キナプリル
14 薬価基準収載医薬品コード	薬価基準収載医薬品に付けられる厚生労働省の分類コード	mhlw_drug_cd	char(24)	2144010F2025
15 個別12桁コード(YJコード)	統一名収載品目の個々の商品を識別するために付与されたコード	yj_cd	char(24)	2144010F2025
16 レセプト電算コード	レセプト請求時に使用される電算処理用のコード	drug_cd	char(18)	610407143
17 成分名	医薬品を構成している物質の名称	general_name	varchar(4096)	キナプリル塩酸塩
18 ブランド名	医薬品名から剤形及び力価情報を除いた名称	brand_name	varchar(1600)	コナン
19 後発品フラグ	後発品のフラグ(1=該当NULL=非該当)	ge_flg	int	0
20 剤形大分類名		form_cat_lrg_name	char(80)	内用薬
21 剤形中分類名		form_cat_mid_name	char(160)	錠剤
22 剤形小分類名		form_cat_sml_name	varchar(400)	錠剤
23 投与量単位	JMDC定義により投与量を算出した場合の単位	admin_amnt_unit	char(160)	錠

3.3.3.1 レセプト電算コード

レセプトに記載される医薬品コードは JMDC マスターファイル中では「レセプト電算コード」と表記されている。9ケタで最初の数字は「6」であり、次の1ケタは剤型(内服薬 1,注射薬 4 そして外用薬 6), 3~5ケタは薬効分類を表す。3~5ケタの薬効分類は総務省が定める日本標準商品分類に基づき,最初の1ケタは大分類,2ケタ目は中分類そして3ケタで小分類という階層構造になっている。

例) 625抗ウイルス薬(作用部位, 目的, 薬効)

62 化学療法剤(成分又は作用部位)

6 病原生物に対する医薬品(用途)

日本標準商品分類は,全ての商品に総務省が分類番号を付与しているもので,総務省統計局サイトで検索できる。医薬品はそのうち上2ケタが「87」である部分である[9]。87396というコードは医薬品(87)→代謝性医薬品(3)→その他代謝性医薬品(9)→糖尿病薬(6)とたどれば糖尿病薬であることがわかる。

レセプト電算コードで困るのは,一部医薬品の薬効分類に 040~047 という全く異なる番号が振られている点である。たとえば抗ウイルス薬のタミフルの医薬品コードは 610443074 と,薬効分類が本来の 625 ではなく 044 となっている。

3.3.3.2 薬価基準(YJ)コード

同一製品につき,レセプト医薬品コードの他に薬価基準(YJ)コードという英字を含む12ケタの別のコードも付与されている。JMDC マスターには15列目に記録されている。このコードは最初の4ケタ(細分類まで)が薬効分類であり,レセプト電算コード

- 第1 レベル:解剖学的部位に基づいた分類(アルファベット1文字)。
- 第2 レベル:治療法メイングループによる分類(2 数字)。
- 第3 レベル:治療法・薬学サブグループによる分類(アルファベット1文字)。
- 第4 レベル:化学・治療法・薬学サブグループによる分類(アルファベット1文字)。
- 第5 レベル:化学構造サブグループによる分類(2 数字)。

これによりたとえばアスピリンは B01AC06 のように表記される。

4. MySQL によるデータ分析の実例

Linux OS の専用サーバーに MySQL を装着し、JMDC データ(csv 形式)を読み取った。処理速度を早めるためレセプト ID 等検索で用いられるフィールドに index を設定した。

分析の例として以下の課題をとりあげる。

レセプトの医薬品データと医薬品マスターとを医薬品コードで結合し、薬剤名、薬効分類(大・中・小・細)別、レセプト種類別そして診療年月別の数量を集計し、Excel のピボットテーブルで操作可能とする

MySQL で集計して CUBE を作成し、それを Excel のピボットテーブルで処理できるようにする^[1]。データウェアハウスでは、計算の対象をメジャー、集計する次元をディメンジョンと呼ぶ。MySQL はビッグデータの分析に適しているが、Excel のピボットテーブルでは MySQL を操作することはできないため、MySQL で抽出した後、ACCESS に読み込み、薬効分類テーブルと結合させた後、Excel の PowerPivot と呼ばれる ACCESS や SQL Server を操作できる機能を使用することとなる。

4.1 医薬品レセプトデータと医薬品マスターとの結合

医薬品レセプトデータのテーブル名は iy、医薬品マスターは iymaster である。両者をリンクするキーは通常レセプトではレセプト電算コードであるが、JMDC データでは JMDC が独自にわりあてた医薬品コード(phma_prod_cd)で両者を連結する。

```
iy inner join iymaster on iy.phma_prod_cd=iymaster/phma_prod_cd
```

4.2 ディメンジョンによる集計

目的とするメジャーは数量(admin_amnt)の合計であり、SQL 上 SUM(admin_amnt)で集計される。ディメンジョンは、薬効分類(4 ケタ)、レセプト種別(claim_type)そして診療年月(YM)の3つであり、これら3つのディメンジョンでメジャーである数量を group by する。なお薬効分類は薬価基準(YJ)コード(yj_cd)の上4ケタであり、4ケタは大→中→小→細分類となっている。最終的に133103行のキューブが得られる。

```
Select drug_name, LEFT(yj_cd,4), YM, claim_type, count(*) from iy inner join
iymaster on iy.phma_prod_cd=iymaster.phma_prod_cd group by drug_name,
LEFT(yj_cd,4), YM, claim_type
```

4.3 ACCESS での結合

薬効分類は薬価基準コードの左 4 ケタで大・中・小・細に分かれる。データウェアハウスでドリルアップ・ダウンできるようにするためには、別に薬効分類のテーブルを作成し結合させなければならない。薬効分類テーブルは【図 6】の通りであり、整然と表示されるよう、薬効分類コードを名称の先頭に付加してある。ACCESS の SQL 文は以下の通りである。

```
SELECT drug_name, YM, TYPE, MAX(YAKKO.SAI), MAX(YAKKO.SHO), MAX(YAKKO.CHU),
MAX(YAKKO.DAI), SUM(SURYO)
FROM DRUG INNER JOIN YAKKO ON DRUG.YAKKO=YAKKO.CODE
GROUP BY drug_name, YM, TYPE;
```

【図 6】 薬効分類テーブル

GOCD	SAI	SHO	CHU	DAI
1111	1111 炭化水素製剤;シクロプロ	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1112	1112ハロゲン炭化水素製剤;	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1113	1113ハロゲンアルコール製剤	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1114	1114エーテル系製剤	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1115	1115溶性バルビツール酸系及	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1116	1116亜酸化窒素製剤	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1119	1119その他の全身麻酔剤	111 全身麻酔剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1121	1121有機ブロム化合物製剤;	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1122	1122メプロバメート系製剤	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1123	1123抱水クロラル系製剤	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1124	1124ベンゾジアゼピン系製剤	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1125	1125バルビツール酸系及びチ	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1126	1126ブロム塩製剤;臭化カリウ	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1127	1127トウキ、センキュウ製剤	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1129	1129その他の催眠鎮静剤、拮	112 催眠鎮静剤、抗不安	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1131	1131フェナセド系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1132	1132ヒダントイン系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1133	1133オキサゾリジン系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1134	1134バルビツール酸系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1135	1135プリミドン系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1136	1136アミノ酪酸系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1137	1137スルホンアミド系製剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1139	1139その他の抗てんかん剤	113 抗てんかん薬	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1141	1141アニリン系製剤;メフェナム	114 解熱鎮痛消炎剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1142	1142ベンジダミン製剤	114 解熱鎮痛消炎剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1143	1143サリチル酸系製剤;アスピ	114 解熱鎮痛消炎剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1144	1144ピラゾロン系製剤;アンチ	114 解熱鎮痛消炎剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬
1145	1145インドメタシン製剤	114 解熱鎮痛消炎剤	11 中枢神経用剤	1 神経系及び感覚器官用医薬

4.4 PowerPivot による Excel への読み込み

Excel は PowerPivot という機能により ACCESS や SQL Server のデータを直に読み込むことができる。Excel のピボットテーブルはスライサーのようなツールを活用す

ることができ、【図 7】は完成した画面であるが、レセプト種別や薬効大分類、中分類でデータをスライスすることが可能となる。

【図 7】ピボットテーブルの表示画面

TYPE	合計 / 数量	列ラベル	入院	入院外	総計
DPC	2113ゴシソニエリキシル0.05mg/mL	DPC		171	171
入院	2113ゴシソニエリキシル0.25mg		468	429	21052
入院外	2113ハーゴキシルKY錠0.125		10774	7512	186580
2個々の器官系用医薬品	2115アブニシオン静注15mg		72	5	77
1神経系及び感覚器...	2115アミノフィリン静注250mg「日医		71	15	37
3代謝性医薬品	2115アンナカ「ホエイ」		5		108
4組織細胞機能用医薬品	2115カフェイン「ケンエー」				8
5生薬及び漢方処方...	2115ネオフィリン原末				3
6病原生物に対する...	2115ネオフィリン錠100mg		42	488	124
7治療を主目的とし...	2115レスピア静注・経口液60mg		1788	62	1850
8麻薬	2115安息香酸Naカフェイン注200mg「		22		22
薬効中分類	2119アカルディカプセル2.5			1078	1078
21値環器官用剤	2119イノバン注0.6%シリンジ		177	10	187
22呼吸器器官用剤	2119カルグート細粒5%		4	7	3
23消化器器官用薬	2119カルグート錠10		2656	1167	905
24ホルモン剤...	2119カルグート錠5		648	688	11173
25泌尿生殖器官...	2119ドパミン塩酸塩0.1%200mLキッ		36	28	64
26外用薬	2119ドパミン塩酸塩点滴静注100mg「		473	7	480
27歯科口腔用剤	2119ドパミン塩酸塩点滴静注100mg「		893	998	1890
29個々の器官系...	2119ドパミン塩酸塩点滴静注200mg「		4		4
11中枢神経用剤	2119ドパミン塩酸塩点滴静注200mg「		548	196	1
12末梢神経用剤	2119ドパミン塩酸塩点滴静注200mg「		8		8
13感覚器器官用薬	2119ピモベンダン錠1.25mg「TE」		33948	16840	218334
31ビタミン剤	2119プロタノールS錠15mg		542	2505	8858
32滋養強壮剤	2119ミルリノン注射液10mg「F」		192	15	208
33血液・体液用薬	2119ユビデカレノン錠10mg「トーフ」		1119	1361	11673
	2119ユビデカレノン錠10mg「トーフ」		477	446	747

4.5 python プログラムによる高度な分析例

python プログラムに SQL を組み込むことにより、統計分析処理等より高度な分析が可能となる。【図 8】は、患者ファイルと個人 ID で結合し、薬剤処方量に性差があるか Welch の T 検定を行う python プログラムの例である。

【図 8】薬剤処方量を男女別に抽出し T 検定を行う python プログラム

```
import pymysql.cursors
from scipy import stats
import numpy

# for PyMySQL
db=pymysql.connect(host="192.168.11.13", user="mysqluser", password="●●●●",
cursorclass=pymysql.cursors.DictCursor)
cursor=db.cursor()
cursor.execute("USE JMDC")
db.commit()

with open('drug_name.txt') as f:
    lines = f.readlines()

for line in lines:
    drug_name = line.strip('¥r').strip('¥n')
    print( drug_name, end=',')

    # 男性の処方量の読み出し。
    sql=("select admin_amnt from iyrspview where drug_name='{0}' and sex=' 男性
';".format(drug_name))
```

```

#print( sql )
cursor.execute(sql)
if cursor != None:
    male_amnt=[]
    for row in cursor:
        male_amnt.append(row['admin_amnt'])
db.commit()

# 女性の処方量の読み出し。
sql=("select admin_amnt from iyrsptview where drug_name='{0}' and sex=' 女性
";".format(drug_name))
#print( sql )
cursor.execute(sql)
if cursor != None:
    female_amnt=[]
    for row in cursor:
        female_amnt.append(row['admin_amnt'])
db.commit()
# Welch の t 検定
t,p= stats.ttest_ind( numpy.array(male_amnt),numpy.array(female_amnt),
equal_var=False );
print( "{0},{1}".format(t,p) )
db.close()
cursor.close()

```

5. 結語

JMDC データは実際のレセプトデータを匿名加工したものであり、レセプトの構造の理解と、医療ビッグデータを分析する技能を習得するための教材としてきわめて有益なものである。今回は医療福祉経営学科の専門科目「医療事務総論」の中で、MySQL でデータを抽出、加工して CUBE 化し、それを Excel のピボットテーブルで操作可能なデータウェアハウス化を行った。

レセプトデータはデータ量が膨大であるだけでなく、多数のマスターファイルとの結合が必要になる等、構造が複雑であり、かつその理解には医療保険制度の専門知識も必要となる。レセプトデータは、地域医療の実態を把握する重要なデータ源として、その重要性は高まっており、計画されている大学院地域情報学研究科の文理融合科目「地域医療情報システム論」の重要な教材となることが期待される。

参考文献

- 1 統合イノベーション戦略推進会議. AI 戦略 2019～人・産業・地域・政府全てに AI [https://www.mext.go.jp/b_menu/shingi/chousa/shotou/056_01/shiryo/attach/_icsFiles/afieldfile/2019/08/30/1420734_002.pdf]
- 2 全国保険者番号簿 2014 年 6 月版. 株式会社サンライズ 2014 年.
- 3 岡本悦司, 神谷達夫. 地方創生データウェアハウス JapanReview.Com の構築とその活用可能性. 福知山公立大学研究紀要 1 巻 1 号 15～30 頁(2017 年).
- 4 厚生省. 疾病, 傷害及び死因統計分類の概要・分類表. 厚生統計協会 1979 年.

-
- 5 疾病，傷害および死因統計分類提要(ICD-10 準拠).厚生省統計情報部 1993 年
 - 6 レセプト電算処理システムコードブック(平成 16 年 4 月版). 社会保険研究所.
 - 7 診療点数早見表 2020 年 4 月版. 医学通信社 2020 年
 - 8 医療情報システム開発センター. ICD・9・CM 手術及び処置の分類. じほう 2001 年
 - 9 総務省. 標準商品分類表[https://www.soumu.go.jp/main_content/000294493.pdf]
 - 10 医療情報システム開発センター. 医療 IT 化と医薬品コード. じほう. 2000 年 11 月
 - 11 長澤亨，原臣司. データベース技術. 医療情報第 2 版・情報処理技術編. 日本医療情報学会 137～173 頁(2016 年)