

データウェアハウス化手法によるレセプトデータの 日韓比較可能性の検討

Comparability of health insurance claims data between Japan and Korea
an analysis using data warehousing technique

岡本悦司 1), 南商堯 2)

1) 福知山公立大学医療福祉経営学科, 2) 柳韓大学 Ubiquitous 保健医療行政学科

要旨

日韓両国の公開されたレセプトデータをデータウェアハウスに加工し、性・年齢階級、傷病分類といった多次元分析を可能にした。傷病分類は、韓国の傷病分類(KCD-7)は世界共通の傷病分類(ICD10)に準拠していたのに対して、日本のデータは 119 分類という独特な分類法を採用していることからそのままでは比較できないが、ICD10 を 119 分類に翻訳することによって、韓国データを日本の傷病分類に近づけることができた。こうしたデータ加工により同一傷病、同一年齢階級間で、たとえば入院日数の割合を比較できることを示した。

キーワード: データウェアハウス, レセプト, 国際疾病分類(ICD), 業務データ

Keywords: data warehouse, health insurance claims, international classification of diseases (ICD), administrative data

1. はじめに

医療受診状況を全数かつ正確に把握するビッグデータとして医療費請求書(日本では慣用的に「レセプト」と呼ばれるので以下「レセプト」と記載する。英語は claims)の活用が注目されている。レセプトのオンライン請求が一般化したことでレセプトデータが毎月収集されるようになり、また億単位のビッグデータを処理できる IT 技術が進歩した。しかしデータは電子化されても、その「次元」によってはたとえば性・年齢階級は比較が容易だが、傷病分類等は国によって独自の分類を採用している場合はそのままのデータでは同一疾病間の比較等が困難な場合がある。本プロジェクトは、データウェアハウス化という、ビッグデータを同次元間で容易に比較できるデータ加工を行い、傷病分類のように国によって分類法が異なる場合には、データ翻訳という加工を施すことによって、日韓両国間データの比較可能性を検討する。

2. レセプトデータをめぐる両国の歴史

日本では 1984 年頃にレセプト請求を紙媒体ではなく、電子化しようという構想が生まれ「レインボー計画」という呼称で一部地域でパイロットスタディが実施されたが、コンピューターによるレセプト審査に対する医療側の反発から頓挫した[1]。

一方、韓国は 2000 年に、それまでは日本のように分立していた医療保険制度が統合されたことをきっかけにレセプト請求の電子化と、医療費請求内容を審査するのみならず医療の質を評価することを法定された健康保険審査評価院 (Health Insurance Review and Assessment, HIRA. 略して審評院[シンピョウイン])が設置され、全国の医療機関から提出されたレセプトデータをデータベース化して医学研究や医療の質評価に活用しはじめた。

日本は 2001 年に森内閣の下で新 IT 戦略(e-Japan)構想が打ち出され「2005 年までに IT 先進国になる」という国家目標の一環としてレセプト電子化目標がかかげられた。しかしながら厚生労働省主導の下でレセプト電子化は遅々として進まず 2005 年時点における電子化率はきわめて低調であり、IT 先進国どころか後進国になりかねない危機感を当時の内閣 IT 戦略室(小泉政権)は抱いた。そこで官邸主導で、韓国をモデルにレセプト電子化と情報活用が推進された。

筆者らは、2006 年当時、総務省の委託研究で韓国におけるレセプト電子化と HIRA データベースの活用状況を調査報告し[2]、それで得られた知見は 2008 年の医療構造改革にも反映された[3]。2008 年医療構造改革は、老人保健制度に変わる独立した後期高齢者医療制度の創設、社会保険庁からの全国健康保険協会の分離独立、特定健診・保健指導(いわゆるメタボ健診)の開始、医療費適正化計画とその「策定・実施・評価」を目的とするナショナルデータベース(NDB)の構築等、日本の医療制度の抜本改革をもたらしたものであり、医療に関する統計も大きく改革された。

現在、医療に関する統計の多くは 2008 年を「元年」とするものが多く、またデータの e-stat 等を介しての提供も大きく進んだ。筆者らは目下、各種医療統計のデータウェアハウス化ととりくんでいるが、それが可能となったのも 2008 年を契機に各種医療統計が整備され、Excel や csv といった加工可能な形式で提供されることが大きい。

「元年」から満 10 年が経過し、医療の最も重要なデータであるレセプト統計について、データウェアハウス(DWH)という手法を用いて、その比較可能性を検討する。

3. 分析手法

日韓両国の入手可能(公開された)レセプト統計をデータウェアハウス化し、両国間で共通に比較できる次元による分析を試みた。

1) 次元

データウェアハウスで「次元(dimension)」とは、数値データを分類する項目を指す。人に関する統計なら、性、年齢階級は比較可能な次元である。医療統計では、傷病分類も重要な次元となる。しかし傷病分類という次元が共通してあっても、もし分類が異なっていると比較可能にならないこともある。傷病分類については WHO が国際疾病分類(international classification of diseases, ICD)を定めて加盟各国に死因や傷病統計に共通して使用するよう呼びかけているのでもし両国のデータが ICD(現在は第10版 ICD10 が使われている)に準拠していれば比較可能となる。国や地域も次元であり、日本は都道府県、韓国は市道と呼ばれるが、国、地域の次元があれば、日本と韓国の比較だけでなく東京都とソウル特別市の比較も可能となる。

2) 階層

次元に、大→中→小関係の階層構造があれば、データウェアハウスの有用な分析手法であるドリルダウン・アップが可能となる。たとえば年齢階級は5歳階級があれば10歳階級にドリルアップが可能であり、もし各歳別データがあれば任意の年齢階級(たとえば7~12歳)にドリルアップすることも可能となる。

地域という次元は、日本では

都道府県>医療圏(又は保健所管轄区域)>市町村>市区町村

という階層関係があり、ドリルダウン・アップが可能である。この場合、下の階層は必ず上の階層に含まれていなければならない。医療圏とは各都道府県の医療計画に指定された複数の市町村を束ねた地域であり、決して他県にまたがらないのでドリルアップ・ダウンが可能である。保健所管轄区域も医療圏とは異なる区域だが、他県にまたがらないので同様である(それに対して税務署管轄区域は一つの市町村内に複数の税務署があつたりするためドリルアップ・ダウンできない)。政令市は、たとえば京都市左京区のように区を有しているので市区町村という階層は市町村の下にくる(=ドリルダウン)。

ICD10 はアルファベットと3桁数字を合わせた4桁が基本分類となり、桁数に応じて、

大分類(アルファベット)>中分類(2桁)>小分類(3桁)>基本分類(4桁)

と階層化されている。具体的には

感染症及び寄生虫症(A)>腸管感染症(A0)>コレラ(A00)>エルトール菌によるコレラ(A00.1)

となっており、もし傷病コード=A00.1なら、SQL関数を用いて表すとLEFT(傷病コード,3)とすれば小分類、LEFT(傷病コード,2)は中分類、LEFT(傷病コード,1)は大分類、とドリルダウン・アップが容易である。

韓国は ICD10 をもとにした韓国標準疾病・死因分類 (Korean standard

classification of diseases、KCD) を使われており現在は KCD-7 が使われている。国民健康保険公団データもその傷病コードはおおむね ICD10 コードに相当している。しかしながら日本のレセプト統計ではいわゆる「119 分類」という独特な分類法があり、医療給付実態調査も全国健康保険協会データも全てこの分類によっている。これは、ICD10 のような医学的体系的な分類ではなく、レセプト分析上の便宜を考えた分類法で、たとえば「0101 腸管感染症」は A0(中分類)に相当するが「0201 胃の悪性新生物」は C16(小分類)に相当する、というふうに体系だっていない。それゆえ、日本の「119 分類」と ICD10 準拠統計とを比較するには、119 分類と ICD10 の「対応表」によって ICD10 を「119 分類」に翻訳することが必要となる。

4. 使用したデータ

日本のデータは全国健康保険協会が公表・提供するレセプトデータであり、韓国は国民健康保険公団が公表するオープンデータを用いた。【表 1,2,3】。

【表1】 両国データの比較

	日本全国健康保険協会データ	韓国国民健康保険公団オープンデータ
対象	中小企業の勤労者と家族	全国民
標本	全数	レセプト有の者100万人を無作為抽出
期間	診療月単位	診療月単位
年	2010～17年度	2016年
年齢階級	10歳階級(75歳未満)	5歳階級
診療種別	入院, 入院外, 調剤, 歯科	入院, 外来, 調剤
地域	47都道府県(事業所所在地)	17市道(8市9道)
診療科	なし	34診療科
傷病分類	119分類	ICD10
データ型	件数, 日数, 費用(調剤含む)	件数, 日数, 費用(給付費+患者負担額)

全国健康保険協会(以下、協会けんぽ)データは 2010 年 4 月～2018 年 3 月の 8 年間分が csv ファイルで提供されており、個票データではなく集計データである(ファイルサイズは約 830MB)。協会けんぽデータで特筆すべきなのは、分母となる被保険者数も、診療月毎、都道府県別かつ性・年齢階級別に提供されている点である。すなわち被保険者数で除することにより「率」を求めることができる。

【表2】日本の全国健康保険協会データの概要

診療月	外来			入院		
	件数	日数	点数	件数	日数	点数
201601	17944061	25398550	20635034651	292516	3040729	14834686726
201602	20508893	29761249	23214783770	307880	3047847	15129310433
201603	21252750	31067059	24294377038	320976	3181885	15823149522
201604	18753773	27413204	21025900383	295258	2905227	14124115350
201605	18772967	26911439	20772167404	304171	3062329	14720304686
201606	19109771	28052243	21724096546	319550	3108392	15612753715
201607	19057623	27947600	21406959148	324429	3209773	15648651766
201608	18236053	26368832	20855359220	334190	3246922	16271504680
201609	18219358	26553450	20866683048	313671	3067205	15054524856
201610	19655702	28887957	21995898644	323217	3191035	15709860478
201611	19651011	28691510	21917402946	316625	3109851	15541719157
201612	20475552	29566008	22599264853	306940	3076565	15175825755
計	231637514	336619101	261307927651	3759423	37247760	183646407124

【表3】韓国国民健康保険公団データの概要

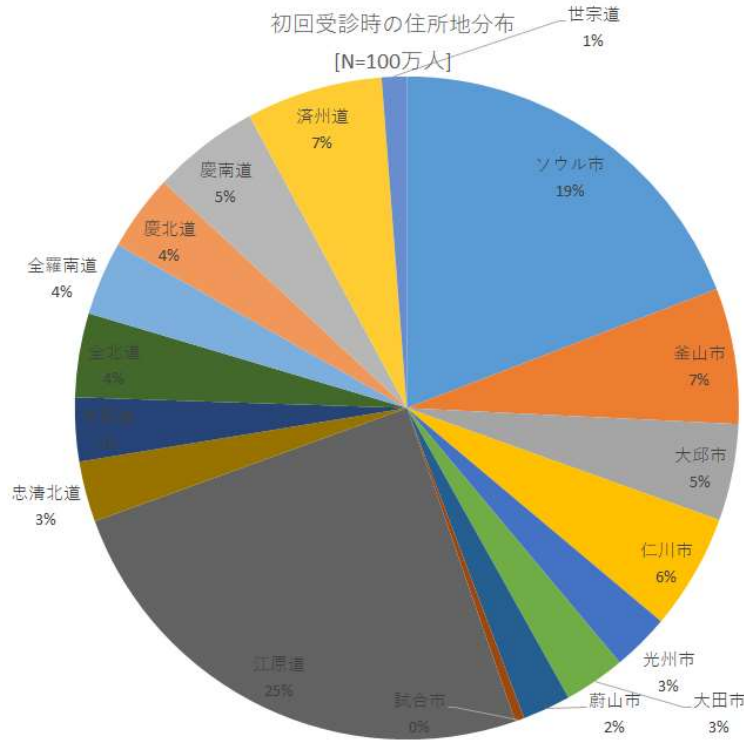
診療年月	件数	給付費	患者負担額	受診日数	入院日数	処方箋日数
201601	1129556	17292074530	60981513590	1389539	2135837	12101876
201602	1152655	16226778680	53395670070	1376159	2094418	12203804
201603	1247897	17646960970	59884133780	1484268	2268665	13242500
201604	1183301	16658035140	55948020110	1406534	2110462	12454098
201605	1154302	17085394380	57645940840	1388339	2145511	12652565
201606	1083942	16539719420	56057060160	1310399	2051141	12294843
201607	1058009	16743166640	57492749820	1295381	2027462	12195790
201608	1088756	17295149600	58452340260	1331243	2113926	12757540
201609	1097128	16316472480	55070106950	1321574	2065735	12697244
201610	1151567	17017368970	58268628210	1383825	2113150	12695028
201611	1199680	17227198280	58650900960	1428562	2175944	13071952
201612	1392183	18398996610	60198405980	1614756	2374416	14190571
計	13938976	204447315700	692045470730	16730579	25676667	152557811

韓国国民健康保険公団データは、全国民から抽出された 100 万人の 1393 万 8976 件の個票データである(csv ファイルでサイズは 1.12GB)。日本の協会けんぽデータと異なる点は、この 100 万人は 2016 年中に一回でも受診した(=レセプトのある)人数であり、1 年間を通して無受診の者いることを考えると、受診していない者も含む全国民より無作為抽出された標本とはいえない、ということ。したがって日本の協会けんぽ

データのように分母にあたるデータがなく「率」を算出することができない(この点、日本のナショナルデータベースも同様)。ただ、両データの性・年齢階級別分布は【表4】の通りでおおむね国民の性・年齢階級構成を反映している。また市道は、同一人でも年間に転居等により変わっている場合があるが、最初の受診時における住所地を市道とした【グラフ】。

【表4】両国データ対象人口の性・年齢階級分布

韓国 抽出された100万人				日本, 全国健康保険協会の被保険者			
年齢階級	男	女	計	年齢階級	男	女	計
00～04歳	24039	22700	46739	00～09歳	2087045	1983557	4070602
05～09歳	25632	24190	49822	10～19歳	2266646	2180747	4447393
10～14歳	24091	22238	46329	20～29歳	2508224	2590738	5098962
15～19歳	29480	28221	57701	30～39歳	3119939	3062156	6182095
20～24歳	29231	32145	61376	40～49歳	3494314	3538710	7033024
25～29歳	29778	31522	61300	50～59歳	2619509	2943673	5563182
30～34歳	32440	35420	67860	60～69歳	2258634	2161317	4419951
35～39歳	36619	39775	76394	70～74歳	324536	371563	696099
40～44歳	37666	40043	77709	75～79歳			
45～49歳	40425	44016	84441	80～84歳			
50～54歳	38141	41594	79735	85～歳			
55～59歳	39863	43369	83232	計	18678847	18832461	37511308
60～64歳	30489	33710	64199				
65～69歳	21926	24061	45987				
70～74歳	16481	19980	36461				
75～79歳	12534	17245	29779				
80～84歳	6743	11900	18643				
85～歳	3347	8946	12293				
計	478925	521075	1000000				



5. 方法

両データに共通し比較可能な次元を抽出し比較可能なかたちに処理した。共通する次元は

- 性別
- 年齢階級・・・全国健康保険協会データは10歳階級なので韓国データは5歳階級を10歳階級に束ねる必要あり(また全国健康保険協会は75歳未満)
- データ型・・・人数, 件数, 日数, 費用(点数)
- 診療種別・・・入院, 外来, 調剤
- 傷病分類・・・主傷病としてふられたコードを使用した。全国健康保険協会データは119分類なので, 韓国データのICD10基本分類(4桁)の上3桁(小分類)を用いて対応表[4]により119分類に翻訳した。なお韓国データの主傷病コードには1桁だけのコード(たとえば, コレラ A00 ではなく, 感染症及び寄生虫症 A といった大雑把なコード)も含まれており(1254万347件中888764件(約7%)), それらは119分類への翻訳はできず, 分類不能とした。

6. 結果

119分類で突合した両国データのうち比較が容易な入院及び外来の日数についてまず傷病大分類別の比較を行った【表5】。

【表5】日韓の受診日数の傷病大分類別比較

	韓国		日本	
	療養日数	入院日数	外来日数	入院日数
韓国：分類不能，日本：病名なし	1793580	5502632	5327390	746541
01感染症及び寄生虫症	421012	508825	15251087	763717
02新生物	422978	771581	11172552	6482051
03血液及び造血器の疾患免疫機構の障害	22620	38187	2001044	347391
04内分泌，栄養及び代謝疾患	587541	995672	23740721	1105600
05精神及び行動の障害			14456228	4371003
06神経系の疾患	544458	968496	7484517	2357512
07眼及び付属器の疾患	676424	766840	21632985	435587
08耳及び乳様突起の疾患	378646	412576	7121612	186641
09循環器系の疾患	2057960	3476549	34143706	4569904
10呼吸器系の疾患	3973383	4315248	73852070	1965274
11消化器系の疾患	838039	1046571	17228804	2461593
12皮膚及び皮下組織の疾患	670671	736821	28156134	373292
13筋骨格系及び結合組織の疾患	2329245	2891793	35085748	2398745
14腎尿路生殖器系の疾患	566183	1374141	14690298	1359615
15妊娠，分娩及び産じょく			1756441	2317991
16周産期に発生した病態	18178	21482	483319	1133777
17先天奇形，変形及び染色体異常	6711	12018	1253399	595971
18症状，徴候及び異常臨床所見・異常検査所見	385407	555561	6119976	400696
19損傷，中毒及びその他の外因の影響	1037543	1281674	15661063	2874846

両データで標本サイズが異なるため，傷病大分類別の日数の割合を観察したところ，両データで大きな違いがみられる分類があった。たとえば，10呼吸器系疾患による入院日数は，日本の協会けんぽデータでの割合は大きくないが，韓国ではかなり大きくなっている。そこで呼吸器疾患についてドリルダウンし，119分類別日数割合を比較した【表6】。

【表 6】119分類による呼吸器疾患入院日数の日韓比較

	全年齢		70～74歳の再掲	
	韓国	日本	韓国	日本
1001急性鼻咽頭炎[かぜ]<感冒>	351346	8284	22690	91
1002急性咽頭炎及び急性扁桃炎	475449	89145	11619	405
1003その他の急性上気道感染症	650099	69164	19438	891
1004肺炎	201500	586336	11079	53611
1005急性気管支炎及び急性細気管支炎	1272611	177269	35976	1270
1006アレルギー性鼻炎	326577	19765	9614	631
1007慢性副鼻腔炎	138222	66594	4505	2492
1008急性又は慢性と明示されない気管支炎	254694	12173	19096	474
1009慢性閉塞性肺疾患	119216	56088	19875	15888
1010喘息	132429	218441	10919	8603
1011その他の呼吸器系の疾患	393105	662015	23960	71248

119 分類別の比較では、日本の入院では 1004 肺炎が圧倒的に多いのに対して、韓国では 1005 急性気管支炎・細気管支炎による入院がきわめて多い。肺炎等の呼吸器疾患による入院は高齢者に多いことから、両データで共通に比較できる 70～74 歳の年齢階級のみ取り出して再掲してみたが傾向は同じであった。

7. 考察

レセプト(医療費請求書)データに含まれるデータ型は、費用(給付費+患者負担額)、日数そして件数とほぼ共通であり、診療に関する次元は、入院と外来、人的次元も、性・年齢階級というように、異なる国であっても共通する部分が大きく、比較可能性が大といえる。医療統計で最も重要な傷病分類についても、WHO が定める国際疾病分類(ICD10)が傷病に関する共通言語として普及しており、韓国の疾病分類(KCD-7)もそれに準拠している。日本のレセプトデータは 119 分類という独自の分類を採用していることから国際比較が困難な場合もあるが、今回分析した韓国の国民健康保険公団のオープンデータは、ICD10 の基本分類まで記録されているため、韓国データを日本独自のデータに「翻訳」することによって両データを少なくとも日本の 119 分類のレベルまでは同一の傷病分類で比較することが可能であった。そうすることによって、たとえば呼吸器疾患を主傷病とする入院日数に大きな違いがあること、119 分類にドリルダウンすると、日本は肺炎による入院日数が多いのに対して韓国は気管支炎による入院日数が肺炎より格段に多い、といった傷病構造の違いを浮き彫りにすることができた。

どの傷病の入院日数やレセプト件数が多いか少ないかは、データ処理上の問題であ

るが、では、日本の呼吸器入院患者は肺炎を主傷病とする者が多いのに対して韓国では気管支炎が多いのか、に対する答えを与えることはできない。それは、両国の医師の診断の違いかもしれないし、あるいは、ひょっとしたら韓国では大気汚染が深刻といわれているのでそうした環境による影響かもしれない[9]。

今回分析したレセプトデータは特定の傷病の多寡はわかってもその原因まで特定することは困難である。しかしながら、月別データを比較することによって、たとえばアレルギー性鼻炎のような季節変動の有無は明らかになる。ちなみに急性気管支炎の主傷病とする入院日数の月間変動の両国の違いは【表7】の通りであり、両国の気管支炎による入院日数に季節変動があるが、その月ごとの変動パターンは両国でかなり異なることはわかる。

【表7】 急性気管支炎による入院日数の季節変動の日韓比較(2016年)

	韓国	日本
201601	111566	15031
201602	126114	10871
201603	126167	10034
201604	117592	10671
201605	102175	12203
201606	73117	11492
201607	62329	14506
201608	61289	13562
201609	83870	18224
201610	107618	25213
201611	128790	19499
201612	171984	15963

8. 結論

レセプトという国際比較の容易なデータを適切に加工しデータウェアハウス化することによって、国間の傷病構造、受療行動そして医療費等を比較可能にできることが示された。性・年齢階級はいうまでもなく傷病分類についてもICD10に準拠したコーディングが適切に行なわれておれば、119分類のような日本独特な分類にも翻訳することができ、月単位、年齢階級別の分析を行うことによって、たとえば大気汚染と呼吸器疾患の入院日数との関連を、因果関係までは明らかにできないまでも、ある程度明らかにすることが示された。

今回用いたデータはいずれも両国でオープンデータとして提供されているものであり、今後さらに詳細なデータを収集して、医療経済や疫学的な研究に活用できるデータウェアハウスに発展させてゆく。

9. 謝辞

本研究は北近畿地域連携推進センター地域研究プロジェクトの助成を受けた。

《参考文献》

- 1 前田征男. レセプト処理の歴史. 医療とコンピューター;5(4):2~5 (1993).
- 2 医療分野における情報化促進のための国内外の実態調査 -レセプトオンライン化に関する韓国実態調査. 総務省委託研究報告書(2006年)
- 3 岡本悦司. レセプトオンライン化はいかに決定されたか -官邸主導の政策決定過程の研究-. ヘルスサイエンス・ヘルスケア 7(2):66-76.
[www.fihs.org/volume7_2/article5.pdf]
- 4 厚生省保険局調査課. 社会保険表章用疾病分類・索引表<119項目>. 社会保険実務研究所(1995).
- 5 【時論】最悪の粒子状物質事態. 中央日報 2019年1月22日
[<https://japanese.joins.com/article/402/249402.html>]