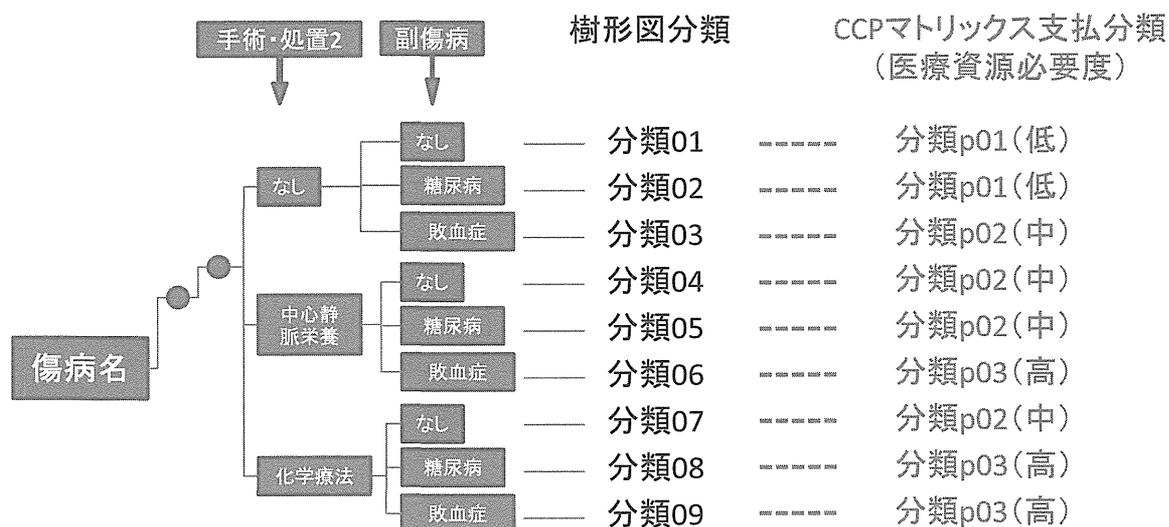


CCPマトリックス分類の考え方(2)

- CCPマトリックスは、分岐条件に基づいて全ての分類を樹形図に書き下した樹形図分類と、医療資源必要度に応じた支払分類との対応表を作成することと同じ意味である。



CCPマトリックスの考え方(3)

- 樹形図分類と支払分類の対応表を用いることにより、枝分かれ構造の制約されずに、どのレベルにおいても分類を集約することが可能となる。
- 例えば、異なる傷病名分類間であっても、医療資源必要度が同等であれば、同じ支払分類とすることができる。
- 傷病名優位のDPC分類の構造を維持しつつ、医療資源必要度に大きな影響を与える手術、処置をより重点的に評価することにより、支払分類の精緻化を図ることができる。
- CCPマトリックス分類対応表を用いることにより、部分的、段階的導入が可能となる。

CCPマトリックスの作成手順

1. DPC定義表の整備

- 診療情報等を含めて医療資源消費に影響する可能性のある項目を追加。
- 必要に応じて、様式1の追加項目の検討
- 必要項目を整理し、体系的にコード化

2. 従来の単純集計と多変量解析等の手法を併用して、集約ロジック素案を設計する。

3. 臨床的判断を加味して、樹形図分類-CCPマトリックス支払分類対応表を作成する。

分類条件に求められること

1. 医療資源必要度を反映するもの

2. DPCデータから情報を得られるもの

- 傷病名情報の精度の向上
- 様式1診療関連情報の検討

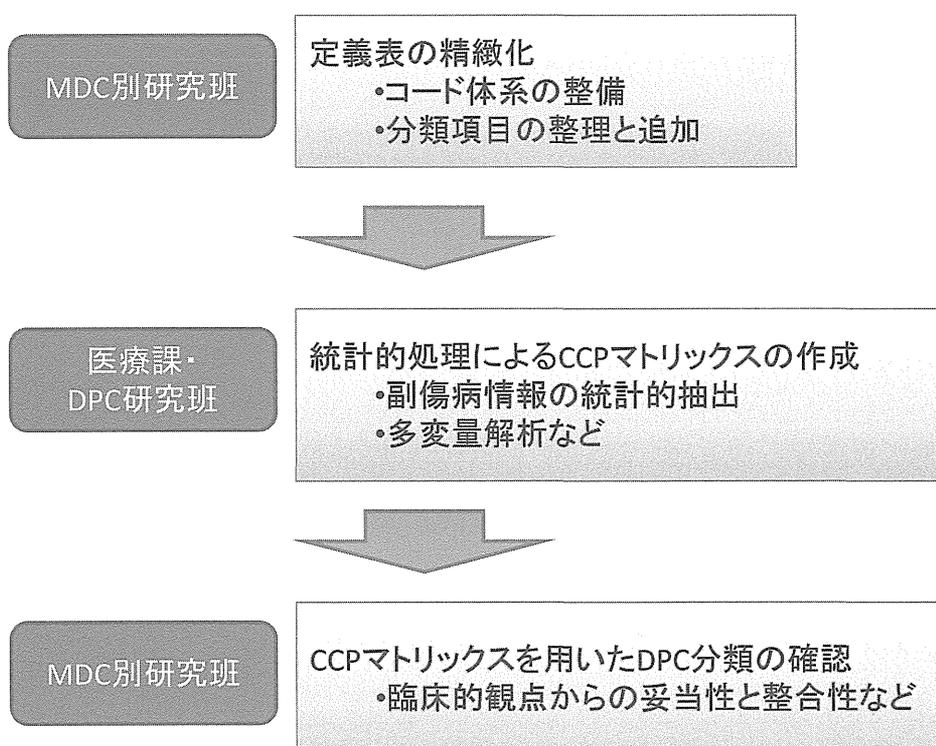
3. 診療を歪める恐れが低いもの

- 過剰な診療行為を誘発しにくいもの
- 治療手技の選択に影響を与えにくいもの
- 必要によりauditの整備

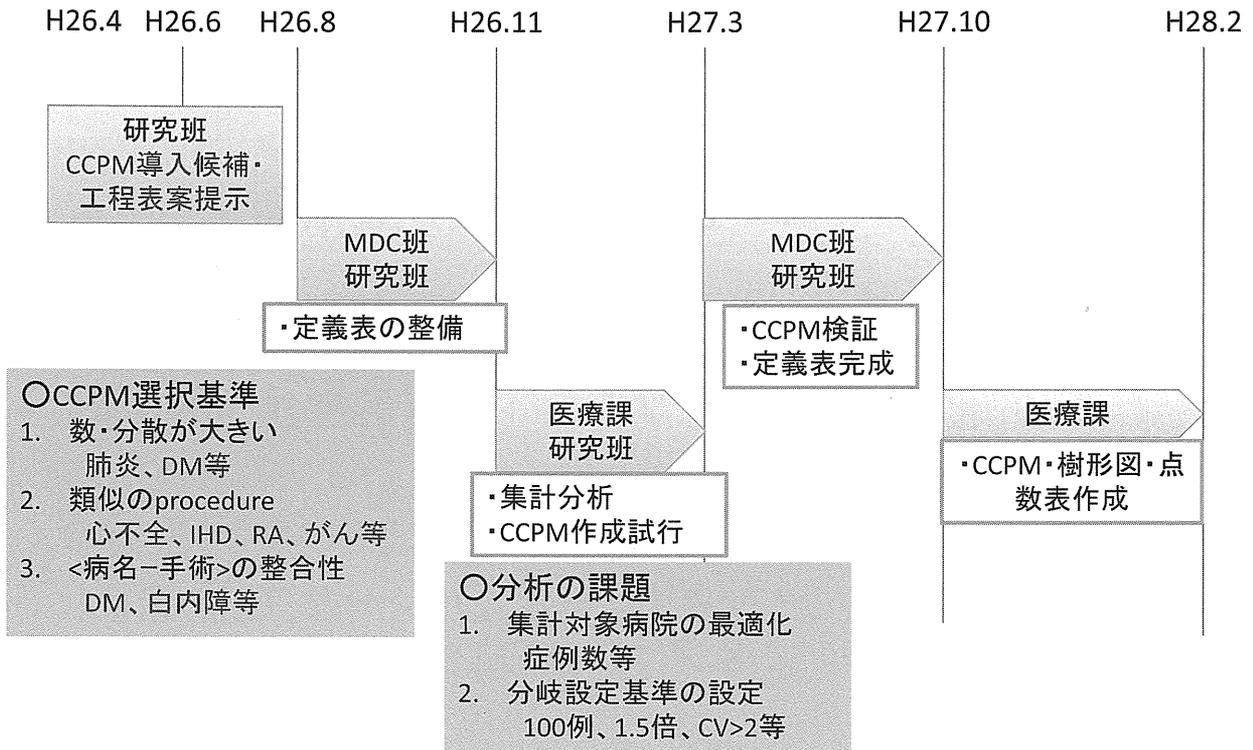
DPC定義表の整備の課題

- 手術、手術・処置等1, 手術・処置等2、副傷病等の体系的な整理が必要
- MDC横断的に統一されたコード体系が望ましい
 - 手術・処置等2では、①リハビリテーション、②中心静脈栄養、③人工呼吸、④血液浄化療法、⑤放射線治療、⑥化学療法、⑦化学療法＋放射線療法、⑧以降明示された高額薬剤など
- 副傷病については、MDC別またはDPC傷病名分類別に整理し、統計的に一定のルールで意味のある副傷病を設定することが可能

DPC分類精緻化の流れ図(案)



CCPM導入の工程表(案)



1. これまでの概要等

前回の「レセプト情報等の提供に関する有識者会議（平成25年6月7日開催）」の概要

- レセプト情報等の提供に関する有識者会議（平成25年6月7日開催）において、厚生労働省科学研究班（伏見班）から、第三者提供を行う上での問題点等について研究結果が報告された。
- 研究結果報告書の中で、個票データについては、「未解決の問題も残っているため、試行的なデータ提供を試みて、課題・問題点を明らかにすることが望ましい」とされている。
- 研究班からの模擬申出・模擬審査を通じて審査基準の明確化等を検討することとなった。



「レセプト情報等の提供に関する有識者会議審査分科会（平成25年9月20日開催）」の概要
○ 研究班（伏見班）より提出された集計表・個票それぞれの模擬申出書について、模擬審査を行い、集計表1件、個票1件、合計2件を承認し、実際にデータの提供を行った。



今回の「レセプト情報等の提供に関する有識者会議（平成26年3月20日）」
○ 研究班（伏見班）から試行的データ提供を通じた検証結果について報告を受け、問題・課題を明らかにし、ガイドラインの策定の方針等を含め、今後のDPCデータ提供の在り方について検討する。

3

2. 伏見班への模擬提供の結果報告

平成25年9月20日の審査分科会における模擬審査に基づき、以下の2本の模擬申し出についてDPCデータの模擬提供を行った。

- ① 集計表データ
産業医科大学 公衆衛生学教室
久保達也・松田晋哉
- ② 個票データ
京都大学 医療経済学分野
猪飼宏・今中雄一

(集計表)

① 久保先生 (産業医大)

DPCデータを用いた若年性胆管がんの記述疫学研究

研究概要

- 印刷事業場での有機溶剤(1,2-ジクロロプロパン)暴露による職業性胆管がんが社会問題となっている。
- ところが今般、問題となっている50歳未満の若年発症胆管がんに関する報告及び情報は未だ限られており、その疫学的特性は不詳である。
- 同特性に関する情報は、臨床医療のみならず労働安全衛生の向上のためにも必要な情報である。
- 本研究では、DPCデータにおける様式1の情報を利用して胆管がんの年齢・地理分布等に関する疫学的記述を行った。

5

方法

- 様式1のみ
- Inclusion Criteria
 - [主病名]、[入院契機病名]、[医療資源病名]、[入院時併存症病名] [入院後発症疾患名]のいずれかに"C221"(肝及び肝内胆管の悪性新生物, 肝内胆管癌)、"C24\$" (その他及び部位不明の胆道の悪性新生物)の症例
- Exclusion Criteria
 - 同一患者の重複を可及的に除外しつつ様式1の臨床情報を効率的に補足するために、「施設コード」「データ識別番号」を連結した文字列(以下、RecIDと呼ぶ)を使用してRecIDが同一の場合は、「入院年月日」が最も早い入院記録のみを解析対象とした

6

申請における研究者側の工夫

- 症例数不足によるマスクを回避するために
 - 50歳未満の症例数ではなく、全症例数と50歳以上症例数を開示請求し、引き算により少数例を把握しようとした。
 - 症例数ではなく割合(%)を開示請求し、全体症例数に割合を掛けることで少数例を把握しようとした
 - 都道府県毎の集計以外に地方別(関東地方等)の集計を請求した
- ダブルカウントの問題を回避するために
 - 施設コード「データ識別番号」を連結した文字列(RecID)を利用して同一施設内の個人を特定した
 - ただし複数施設受診のダブルカウントは回避不能のため研究計画自体の集計単位を個人ではなく入院とすることとした

7

結果1-1: 患者の性・年齢・病期分布

年齢・病期	患者数	うち男性患者数	肝内胆管癌患者数†	胆道癌患者数‡	初発患者数	早期患者割合※
30歳未満	85	43	25	60	17	38.5%
35歳未満	70	34	20	51	23	38.9%
40歳未満	171	105	42	130	79	28.6%
45歳未満	306	185	97	213	132	23.4%
50歳未満	489	308	108	384	226	31.1%
55歳未満	896	590	241	662	485	27.6%
60歳未満	1,696	1,122	441	1,265	951	27.0%
65歳未満	4,260	2,917	1,017	3,267	2,459	31.2%
70歳未満	5,100	3,364	1,138	3,988	2,975	35.2%
75歳未満	6,684	4,373	1,423	5,311	3,896	37.0%
80歳未満	8,015	4,958	1,520	6,541	4,587	37.4%
85歳未満	6,970	3,958	1,061	5,934	3,853	39.6%
85歳以上	6,433	2,730	798	5,656	3,515	40.2%
不明	-	-	-	-	-	-

※: 病期が判明している症例に占める病期Ⅱ以下の患者割合

†: C221が病名コードに含まれた患者(病期毎の集計においては医療資源病名のみで判定)

公表の基準に基づき、10以下のセルはマスクしている
(実際には10以下のセルもふくめ提供)

‡: C24\$が病名コードに含まれた患者(病期毎の集計においては医療資源病名のみで判定)

※: 審査基準の検証を行うため、10未満のセルについても提供を行い、公表時にマスクをしている。(以下同じ。)

ダブルカウントを排除した形でも全国の胆管がん症例が豊富に記述された

8

結果1-2: 患者の病期分布

年齢・病期	患者数	うち男性患者数	肝内胆管癌患者数†	胆道癌患者数‡	初発患者数	早期患者割合※ 50歳未満患者割合
0	63	41	-	37	-	-
I	2,871	1,777	248	2,434	-	1.8%
INOS	-	-	-	-	-	-
IA	98	56	-	67	-	-
IB	76	48	-	64	-	-
IBI	-	-	-	-	-	-
II	3,257	2,035	485	2,499	-	1.5%
II NOS	-	-	-	-	-	-
II A	76	50	-	63	-	-
II B	66	47	-	-	-	-
III	3,271	2,008	571	2,356	-	1.8%
III NOS	13	-	-	-	-	-
III A	59	37	-	21	-	-
III B	60	35	-	40	-	-
III C	-	-	-	-	-	-
IV	485	317	82	210	-	2.3%
IV NOS	778	444	223	474	-	3.0%
IV A	2,844	1,668	547	2,007	-	2.0%
IV B	3,950	2,310	1,414	2,049	-	2.8%
IV C	-	-	-	-	-	-
不明	5,622	3,292	970	4,143	-	2.2%
未入力(NULL)	17,568	10,507	967	2,817	-	3.6%

※: 病期が判明している症例に占める病期II以下の患者割合

†: C221が病名コードに含まれた患者(病期毎の集計においては医療資源病名のみで判定)

公表の基準に基づき、10以下のセルはマスクしている
(実際には10以下のセルもふくめ提供)

‡: C24\$が病名コードに含まれた患者(病期毎の集計においては医療資源病名のみで判定)

✓病期入力状況(不明・未入力が56.3%と過半数)であったが、豊富な病期情報が包括的に示された

9

結果2: 入院時併存・合併症

	N	初発肝内胆管癌患者†に占める割合	初発胆道癌患者‡中に占める割合
B: 50歳未満患者集計			
原疾患			
初発肝内胆管癌患者†	149	-	-
初発胆道癌患者‡	267	-	-
合併症(入院時併存症)			
胆のう胆管及び膵の障害	200	26.85%	59.93%
胆管結石症	11	-	-
胆道のその他の疾患	163	21.48%	49.06%
胆管炎	56	-	-
慢性ウイルス肝炎	13	-	-
慢性C型肝炎ウイルス感染	-	-	-
慢性B型肝炎ウイルス感染	-	-	-
肝硬変	-	-	-
アルコール性肝硬変	-	-	-
肝線維症・肝硬化症	-	-	-
胆汁性肝硬変	-	-	-
胆のう胆管及び肝の先天奇形	14	-	-
胆管の先天奇形	14	-	-
肝吸虫感染	-	-	-
潰瘍性大腸炎	-	-	-
クローン病	-	-	-

※: 病期が判明している症例に占める病期II以下の患者割合

†: C221が医療資源病名にコードされた初発患者のうち当該合併症を持つ患者の割合

公表の基準に基づき、10以下のセルはマスクしている
(実際には10以下のセルもふくめ提供)

‡: C24\$が医療資源病名にコードされた初発患者のうち当該合併症を持つ患者の割合

✓初発患者のみを対象に胆管癌のリスク因子として知られる合併症コードがどの程度発生しているかを集計
✓一般の有病率と比較することでリスク因子の検討が可能となる可能性がある(容易ではないが)
✓少数症例を対象とした集計でも割合で開示請求することでN数を同定可能である

10

結果3-1 患者の地理的分布(地方)

	患者数	男性患者数	女性患者	50歳以上患者数	50歳未満患者数
地方別集計					
北海道地方	2,307	1,389	918	2,267	40
東北地方	3,756	2,185	1,571	3,663	93
関東地方	11,507	7,046	4,461	11,122	385
中部地方	6,786	4,071	2,715	6,615	171
近畿地方	7,241	4,321	2,920	7,022	219
中国地方	2,564	1,542	1,022	2,508	56
四国地方	1,440	844	596	1,414	26
九州沖縄地方	5,427	3,195	2,232	5,300	127
不明	147	94	53	-	-

公表の基準に基づき、10以下のセルはマスクしている
(実際には10以下のセルもふくめ提供)

✓女性患者数と50歳未満患者数は引き算により算出した
✓人口動態調査等と組み合わせることで地域の有病率を検討することが可能と考えられた

11

結果3-2 患者の地理的分布(都道府県)

	患者数	男性患者数	女性患者	50歳以上患者数	50歳未満患者数		患者数	男性患者数	女性患者	50歳以上患者数	50歳未満患者数
北海道	2,307	1,389	918	2,267	40	滋賀県	416	250	166	-	-
青森県	499	295	204	482	17	京都府	809	489	320	788	21
岩手県	647	393	254	633	14	大阪府	2,891	1,796	1,095	2,782	109
宮城県	707	425	282	688	19	兵庫県	1,736	966	770	1,695	41
秋田県	539	296	243	527	12	奈良県	441	257	184	426	15
山形県	535	299	236	524	11	和歌山県	399	240	159	-	-
福島県	829	477	352	809	20	鳥取県	272	165	107	-	-
茨城県	862	518	344	846	16	鳥根県	310	182	128	-	-
栃木県	661	389	272	634	27	岡山県	613	366	247	603	10
群馬県	594	345	249	580	14	広島県	928	558	370	902	26
埼玉県	1,769	1,089	680	1,703	66	山口県	441	271	170	-	-
千葉県	1,534	965	569	1,490	44	徳島県	286	173	113	-	-
東京都	3,494	2,080	1,414	3,364	130	香川県	357	218	139	-	-
神奈川県	2,593	1,660	933	2,505	88	愛媛県	425	236	189	414	11
新潟県	802	485	317	786	16	高知県	372	217	155	-	-
富山県	447	278	169	431	16	福岡県	1,827	1,059	768	1,771	56
石川県	425	227	198	410	15	佐賀県	292	164	128	-	-
福井県	302	178	124	-	-	長崎県	550	320	230	529	21
山梨県	240	145	95	-	-	熊本県	649	391	258	635	14
長野県	975	595	380	956	19	大分県	459	269	190	-	-
岐阜県	665	380	285	645	20	宮崎県	360	213	147	-	-
静岡県	1,189	733	456	1,166	23	鹿児島県	833	501	332	-	-
愛知県	1,741	1,050	691	1,688	53	沖縄県	457	278	179	442	15
三重県	549	323	226	531	18	不明	147	94	53	-	-

公表の基準に基づき、10以下のセルはマスクしている
(実際には10以下のセルもふくめ提供)

✓結果3-1と同様に女性患者数と50歳未満患者数は引き算により算出した
✓N数が豊富なDPCデータにより全都道府県における若年発症胆管がん患者数が記述可能であった

12

(個票データ)②京都大学医療経済学分野

患者居住地二次医療圏単位でみた地域別医療パフォーマンスの分析

心筋梗塞・心不全・脳梗塞の3疾患につき、主要な医療の質指標を患者居住地二次医療圏別に分析した。



13

(個票データ)②京都大学医療経済学分野

臨床指標の例:

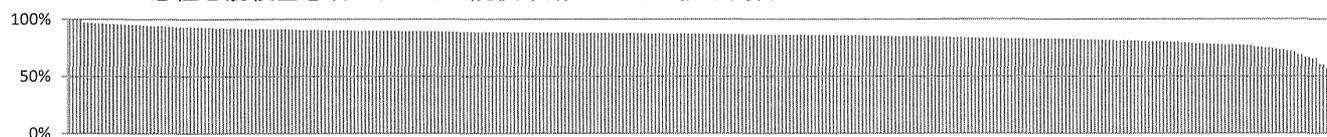
1. 急性心筋梗塞患者における
 - ・入院後早期アスピリン投与割合
 - ・アスピリン投与割合
 - ・βブロッカー投与割合
 - ・ACE阻害剤もしくはアンギオテンシンⅡ受容体阻害剤の投与割合
2. 心不全患者における
 - ・βブロッカー投与の割合
 - ・ACE阻害剤もしくはアンギオテンシンⅡ受容体阻害剤投与の割合

3. 急性心筋梗塞(再発性心筋梗塞含む)患者・慢性心不全患者における
 - ・心臓リハビリ実施割合
4. 脳卒中の診断で入院し、
 - ・リハビリ治療のための評価を受けた症例の割合
 - ・リハビリ治療を受けた症例の割合
 - ・入院後早期にリハビリ治療を受けた症例の割合
5. 脳梗塞(TIA含む)の診断で入院し、
 - ・第2病日までに抗血栓療法を受けた症例の割合
 - ・入院中に未分画ヘパリンを投与されなかった症例の割合

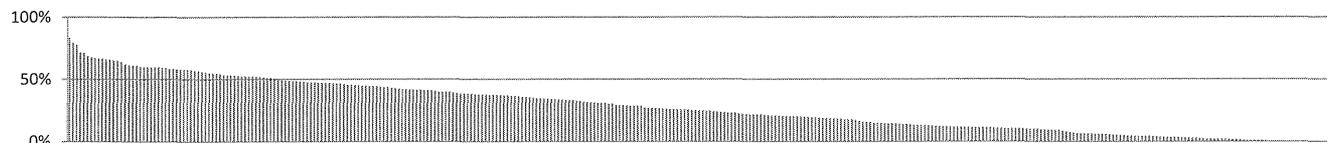
14

(個票データ)②京都大学医療経済学分野

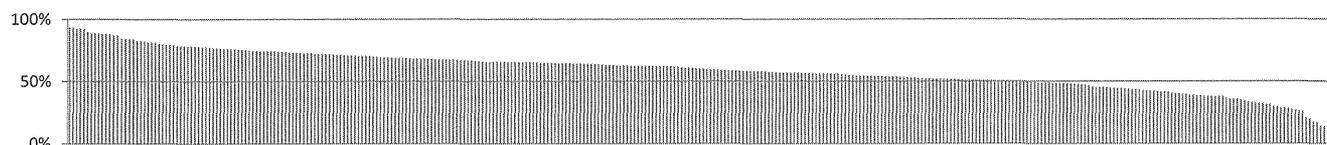
急性心筋梗塞患者における入院後早期アスピリン投与割合



慢性心不全患者・急性心筋梗塞(再発性心筋梗塞含む)患者における心臓リハビリ実施割合



脳梗塞の診断で入院し、入院後早期にリハビリ治療を受けた症例の割合



15

3. 今回の模擬提供を通して

①伏見研究班からの報告

- DPCデータの個人・医療機関特定可能性についての検証
- 審査基準(案)の検証
- セキュリティについて

②セキュリティ監査の報告

16

①伏見研究班からの報告

患者、医療機関等の特定可能性について

- DPCデータを審査する際は、「患者」、「医療機関」、「保険者」、「医師個人」等にとって不利益な情報が公表されないよう、以下の視点から審査が行われたが、他に留意すべき個別事項、共通事項がある。

視点	患者情報	医療機関情報	保険者情報	医師個人情報
留意すべき個別事項	・ 郵便番号 ・ 生年月日 等	・ 毎年公開されている集計表との組み合わせ 等	・ 比較的小規模な保険者 等	・ 医師コード 等
留意すべき共通事項	<ul style="list-style-type: none"> ・ 稀少な疾患 ・ 稀少な術式・処置 ・ 稀少な薬剤・医療材料の使用 等 ・ 報道等で先に患者の個人情報¹が周知された症例 			

17

患者、医療機関等の特定可能性について

・ 患者個人の特定可能性について

実際には行っていないが、新聞等で医療機関や病状が詳細に報道された症例や、著名人で生年月日等個人情報が公開されている人物について、現状のマスク化にもかかわらず複数条件の絞り込みでレコードを同定できうるケースは予想された。

今回の集計表請求では個人の特定は不能であった。ただし粒度の高い集計表を組み合わせることで特定が可能になる可能性は当然ある。

・ 医療機関の特定可能性について

医療圏別に各医療機関を症例規模順で並べると、大規模な施設に限っては、病床規模やDPC評価分科会の退院患者調査との関連から一定程度照合・類推できる可能性があった。

今回の集計表請求では医療機関の特定は不能であった。ただし粒度の高い集計表を組み合わせることで特定が可能になる可能性は当然ある。

・ 医師個人の特定可能性について

手術や手技等で地域で限られた医師のみが行っている状況が学会等で発表された場合、二次医療圏単位での分析では一定程度類推できる可能性があった。

18

提供データの安全性について

【① 全般的な事項】

①-1 [DPCデータ特有の基準]

- ・「DPC導入の影響評価に関する調査（以下「退院患者調査」という。）」の結果報告において毎年公表されている集計内容（MDC別・救急医療入院、診断群分類毎の集計、疾患別・手術別集計等）と重複する申請については、提供は行わない。もし必要とする場合は、その理由が明確に申出書に記載されているか確認したうえで、慎重な審査を行う。
- ・「DPCデータ以外の情報（申出者が保持する情報等）」と結びつけて集計を行う申出は、慎重な審査を行う。
- ・退院患者調査の公表情報以上に患者・医療機関等が特定される恐れがある申出は、慎重な審査を行う。
- ・提供対象医療機関はDPC対象病院までとし、DPC準備病院、出来高算定病院のDPCデータの提供は行わない。

①-2 [NDBデータと同様の基準]

- ・多数の項目を用いた探索的研究を目的とした申出は、慎重な審査を行う。
- ・「複数の研究」が1申出に盛り込まれている場合は、慎重な審査を行う。
- ・研究に際して抽出項目の指定や研究目的との関連については、慎重な審査を行う。
- ・「必要最小限」の範囲で必要な情報を提供することから、具体的な個々のデータ集計方法、解析方法まで分解した記述がない場合は、慎重な審査を行う。

項目の組合せにより症例を同定することは完全には防ぎ得ない。希少疾患や希少な手技などデータ項目の組合せで容易に小さなセルを発生できるケースは、地理情報を粗くする等の工夫を要する。

19

集計表に関する提供データの安全性について

【② 提供に関する審査基準】

- ・生年月日については、年月までとする。
- ・入院年月日・退院年月日については、在院日数・術前日数での提供も可とする。
- ・患者住所地域の郵便番号については、原則提供を行わないこととし、経年データ分析等を行うため必要な場合は、2次医療圏までの提供とする。
- ・施設コード、データ識別番号（患者情報）提供については、原則提供を行わないこととし、提供を行う場合は新たな通し番号を付番する。
- ・保険者番号の提供については、原則提供を行わないこととし、経年データ分析等を行うため必要な場合は、提供することとする。

- 集計表において各データ項目（生年月日等）で粒度を落としても、多様な項目の組み合わせにより症例を絞り込んで同定できることを完全には防ぎえない。
⇒従って、研究の意義を損なうことなく、かつ個人や医療機関の特定ができないようにする普遍的な審査基準を策定することは困難である。
⇒提供データにおける匿名性とは別に、抽出過程で取り扱っているデータのマスキング条件については、第三者の同席や第三者による操作ログのチェックなどの対策を前提に、ある程度緩和できるのではないかと。

集計表に関する提供データの安全性について

【③ 公表に関する審査基準】

- ・生年月日については5歳刻みとし、85歳以上については同一の集計とする。
- ・患者の公表単位は、公表される研究の成果物において最も狭い地域区分の集計単位は2次医療圏とする。
- ・医療機関の所在地の公表単位は研究の成果物において最も狭い地域区分の集計単位を都道府県単位とする。
- ・保険者番号の公表については、4制度（被用者保険、国民健康保険、後期高齢者、共済組合）とする。
- ・**最小集計単位**については、研究の成果物において、1つのセルあたりの集計値が患者の数が10を下回る場合、医療機関の数が5を下回る場合は、その数値をマスクする。

・最少集計単位での匿名化を通じて、最少セルが一定基準を下回る場合にはマスクさせるほかに、周辺のセルと融合させる等の配慮も同等に有効であろう。

・保険者番号の公表については、4制度（被用者保険、国民健康保険、後期高齢者、共済組合）に限らず、研究に応じて独自のグループ分け（数グループ程度）を行っても安全性には影響を与えないのではないかと。

－むしろ、小規模の保険者や公費医療などを単独でラベルしないルールがより実効的。

・すでに医療機関に関する情報は、厚労省の退院患者調査の結果報告で一定程度公表されているので、医療機関の所在地は特定されてもよいのではないかと。

・公表時に匿名化を義務付ける現在の提供方法でよい。提供時に匿名化を行うとすれば解析に大きな制約が加わると懸念される。

一定の症例数制限を設けても、背反的事象や割合の公開請求を行うことでデータは容易に算出されてしまう。10未満をマスクして提供を行ったほうがいいのか、全部提供した後に公表時に利用者がマスクをしたほうがいいのかという議論があるが、上述のことから、公開時に利用者がマスクすることが個人情報保護の観点から適切と思われる。

21

集計表に関するセキュリティの安全性について

【セキュリティの整備】

DPCデータの提供を行うにあたって、利用者に対し一定のセキュリティ水準を具備することを求めることとしている。

○利用にあたっての基本的な条件（案）

- ・提供したレセプト情報等の情報システム等への複写は、前段階でのデータが消去されない限り原則1回のみ。この原則は、厚生労働省から提供されたレセプト情報等の元データだけでなく当該元データから作成される全ての中間生成物も含め適用される。
- ・利用・保管場所は、あらかじめ申し出られた施設可能で入退室管理を行っているスペースのみとし、原則として持ち出されないこと。
- ・レセプト情報等を複写した情報システムはインターネット等の外部ネットワークには接続しないこと。
- ・レセプト情報等は事前に申し出られた利用者以外の者が利用してはならないため、これを担保するための情報システムの認証等の措置も必要。
- ・学部、研究室などの合理的な範囲内でガイドライン等のルールを定めた運用管理規程も必要。
- ・運用管理規程の運用状況を確認するための内部監査（自己点検）規程も必要。

- ①当施設で複数の研究者が出入りする環境でデータ利用を行った場合、NDBデータ専用のサーバー設置スペースの確保、生体認証による入退室記録に200万円（当施設での一例）、情報セキュリティマネジメントの運用自体に数人月を要し、大きな負担を要している。
 - ②受け手がセキュリティを構築・維持するのは大変なので、DPCデータを用いた研究を普及するためには、サンプリングデータでの提供やオンサイトセンターを整備することが望ましいのではないかと。
 - ③データ利用期間終了後も論文査読等でデータの再検証を求められる可能性があるため、利用期間終了時に加工データも併せて返納し、期間限定・あるいは理由を限定したうえで、後日オンサイトセンターにおいて参照できる環境が必要ではないかと。
- 利用者がセキュリティ基準を満たす体制を構築・維持することは労力、コストの観点から非常な負担である。日本全体の研究資源の効率的な活用という視点からも、DPCデータを用いた研究を普及するためにはオンサイトセンターを整備することが望ましい。

22

個票に関する提供データの安全性について

【② 提供に関する審査基準】

- ・生年月日については、年月までとする。
- ・入院年月日・退院年月日については、在院日数・術前日数での提供も可とする。
- ・患者住所地域の郵便番号については、原則提供を行わないこととし、経年データ分析等を行うため必要な場合は、2次医療圏までの提供とする。
- ・施設コード、データ識別番号（患者情報）提供については、原則提供を行わないこととし、提供を行う場合は新たな通し番号を付番する。
- ・保険者番号の提供については、原則提供を行わないこととし、経年データ分析等を行うため必要な場合は、提供することとする。

○ 日付情報に下記の加工を事前に加えることで、同定可能性を下げてはどうか。

- 生年月日は提供せず、代わりに入院時年齢・退院時年齢を提供する。
- 入院年月[日]は原則的に提供せず、代わりにE/Fファイルを含むすべての日付情報を、入院日からの通算日数に変換して提供する。（年月情報は季節変動や経年データ分析等を行うなど特定の研究目的に限って提供する。）
- 患者住所地郵便番号は提供せず、二次医療圏に変換して提供する。
- ただし下記のいずれかの情報は併せて提供しても良いのではないかと。
- 医療機関郵便番号 あるいは
- 住所地市区町村が医療機関と同一であるか
- 自治体等が行政上の必要から分析を行う場合には住所地郵便番号を提供しても良いのではないかと。

○ - 個票において各データ項目（生年月日等）で粒度を落としても、多様な項目の組み合わせにより症例を絞り込んで同定できることを完全には防ぎえない。

⇒従って、「研究の意義を損なうことなく、かつ個人や医療機関の特定ができないようにする普遍的な審査基準を策定することは困難」である。

⇒提供データにおける匿名性とは別に、抽出過程で取り扱っているデータのマスク条件については、第三者の同席や第三者による操作ログのチェックなどの対策を前提に、ある程度緩和できるのではないかと。

23

個票に関する提供データの安全性について

【③ 公表に関する審査基準】

- ・生年月日については5歳刻みとし、85歳以上については同一の集計とする。
- ・患者の公表単位は、公表される研究の成果物において最も狭い地域区分の集計単位は2次医療圏とする。
- ・医療機関の所在地の公表単位は研究の成果物において最も狭い地域区分の集計単位を都道府県単位とする。
- ・保険者番号の公表については、4制度（被用者保険、国民健康保険、後期高齢者、共済組合）とする。
- ・**最小集計単位**については、研究の成果物において、1つのセルあたりの集計値が患者の数が10を下回る場合、医療機関の数が5を下回る場合は、その数値をマスクする。

・ 最少集計単位でのk匿名化を通じて、最少セルが一定基準を下回る場合にはマスクさせるほか、周辺のセルと融合させる等の配慮も同等に有効であろう。

・ 保険者番号の公表については、4制度（被用者保険、国民健康保険、後期高齢者、共済組合）に限らず、研究に応じて独自のグループ分け（数グループ程度）を行っても安全性には影響を与えないのではないかと。

- むしろ、小規模の保険者や公費医療などを単独でラベルしないルールがより実効的。

・ すでに医療機関に関する情報は、厚労省の退院患者調査の結果報告で一定程度公表されているので、医療機関の所在地は特定されてもよいのではないかと。

・ 公表時に匿名化を義務付ける現在の提供方法でよい。提供時に匿名化を行うとすれば解析に大きな制約が加わると懸念される。

24

個票に関するセキュリティの安全性について

【セキュリティの整備】

DPCデータの提供を行うにあたって、利用者に対し一定のセキュリティ水準を具備することを求めることとしている。

○利用にあたっての基本的な条件（案）

- ・提供したレセプト情報等の情報システム等への複写は、前段階でのデータが消去されない限り原則1回のみ。この原則は、厚生労働省から提供されたレセプト情報等の元データだけでなく当該元データから作成される全ての中間生成物も含め適用される。
- ・利用・保管場所は、あらかじめ申し出られた施設可能で入退室管理を行っているスペースのみとし、原則として持ち出されないこと。
- ・レセプト情報等を複写した情報システムはインターネット等の外部ネットワークには接続しないこと。
- ・レセプト情報等は事前に申し出られた利用者以外の者が利用してはならないため、これを担保するための情報システムの認証等の措置も必要。
- ・学部、研究室などの合理的な範囲内でガイドライン等のルールを定めた運用管理規程も必要。
- ・運用管理規程の運用状況を確認するための内部監査（自己点検）規程も必要。

- ①入退室管理を行っているスペースにおいて、関係者以外の入室を制限したり、目隠しを行うなど、窃視防止の対策が必要ではないか。
- ②当施設で複数の研究者が出入りする環境でデータ利用を行った場合、NDBデータ専用のサーバー設置スペースの確保、生体認証による入退室記録に200万円（当施設での一例）、情報セキュリティマネジメントの運用自体に数人月を要し、大きな負担を要している。
- ③受け手がセキュリティを構築・維持するのは大変なので、DPCデータを用いた研究を普及するためには、サンプリングデータでの提供やオンサイトセンターを整備することが望ましいのではないか。
- ④データ利用期間終了後も論文査読等でデータの再検証を求められる可能性があるため、利用期間終了時に加工データも併せて返納し、期間限定・あるいは理由を限定したうえで、後日オンサイトセンターにおいて参照できる環境が必要ではないか。

25

オンサイトセンターに期待される機能

- ・ 生に近いデータを閲覧できる環境
- ・ データサイエンティスト(専任)により、データ特性を踏まえた解析計画のカウンセリング
- ・ 終了したプロジェクトについて、中間テーブル等のファイルを保管し、申請に応じて参照できる環境

※全国で複数拠点の運用が望ましい。

※人員配置は利用実績に応じて定期的に見直す必要がある。

26

オンサイトセンターで配慮すべき セキュリティ上の課題

- 不適切なデータ参照を防ぐため、SQL文のログを残し、担当者が事後チェックできる環境づくり
- 研究者がデータを持ち帰る際に、個人特定可能性が十分に低いことを適切に判断できる体制
- 生に近いデータを安全に保管・運用するための堅牢な情報セキュリティマネジメント
- 研究利用を許可されたものを正確に認証する仕組み

27

②セキュリティ監査の報告

【セキュリティ監査の目的について】

ガイドライン策定にあたり、想定しているセキュリティ項目の他に盛り込む項目があるのか、模擬申出書に記載されたセキュリティが実際に確保されているか、セキュリティ監査を実施する方法等を検証するため、厚生労働省科学研究班（伏見班）のご協力を得て、実際にDPCデータの利用場所、保管場所に職員等を派遣し、模擬的にセキュリティ監査を実施した。

【検査の概要】

- ①日時：平成25年11月15日（金）13時30分～18時10分
- ②場所：京都大学大学院 医療経済学分野
- ③監査実施者：厚生労働省保険局医療課職員2名及び総務課職員1名（※）
※レセプト情報等の提供で既に監査を行っていることから、総務課職員をアドバイザーとして助言を得た。
- ④検査内容
 - 書類に記載されたセキュリティ要件の確認
 - 現場での運用状況の確認

28

セキュリティ監査の概要

【安全管理がなされていた点】

- 厚生労働省から提供された元データはセキュリティ基準が高いサーバ室に保管されており、生体認証による入退室管理がなされていた。研究室へ入室する場合もID認証管理を行っており、登録された職員等の帯同がなければ入室することはできない
- DPCデータの利用は研究室内のサーバ室内のみであった
- DPCデータの分析等の業務は保管サーバから作業用PCへ教室内LANを介さず専用ネットワークを用いて情報を受け渡しを行っている。また、格納サーバに格納したデータは保管サーバから呼び出すことができるが、情報が書き換えられないよう情報の流れを一方向となるよう工夫がされていた
- ウィルス対策ソフトを更新する際はDPCデータを外付けHDDに退避させ教室内LANに接続するなど、インターネットの接続によるウィルス感染を防止する対策をとっていた

【対応が必要と思われる点】

- サーバ室への入退は生体認証により利用者が制限されているが、監視する位置に防犯カメラ、監視カメラは設置されていなかった
- サーバ室への入退において、携帯電話等の情報機器の持ち込みに伴う情報漏洩への対策については明確に定められていなかった。

29

データ提供を行うための留意すべき点

【セキュリティに関すること】

- 生体認証による入退室管理などセキュリティが高いサーバ室等が必要がある
- サーバ室を設置したとしても維持・運用にコストや人的資源がかかり、研究者個人が負担することは難しい

【提供に関すること】

- DPCデータはDPC評価分科会において「DPC導入の影響評価に関する調査」として一定程度のデータを公表しているため、研究者は提供申出をする場合は確認する必要があると考える
- 研究者は研究内容、提供項目等を明確にする必要があるが、どのような研究に有効なデータかを知る必要があると考える
- 個票を提供する場合、匿名化をしたうえで提供を行うが複数の条件で絞り込むことで、ニュースや新聞など公表されている情報から個人が特定される恐れがあると考え
- 匿名化をしたうえで提供するが10未満のセルが多数存在した場合、本来行いたかった研究ができなくなる恐れがあるためある程度制限を設けずに提供することも検討すべきだが、その場合セキュリティ審査を厳しくする必要があると考える

30

II. 分担研究報告

DPC 分析用データセットの作成・開発について

堀口 裕正 国立病院機構本部総合研究センター 主席研究員

研究要旨

本研究班において、収集した DPC データは、データセットの量が大きく、一般的な研究者が保有する分析環境（コンピュータの能力やデータを保管するストレージの量等）では処理が行えない状況となっている。また、その膨大なデータのうち、矛盾するレコードや、研究で使用するには留意が必要なデータも混じっている。

そこで、いくつかのデータ処理を行うことによって、データを分析可能なものに絞り込み、さらに分析に必要な様々な処理を加えてデータセットを作成し、さまざまな研究が実施しやすい環境を構築することを行った。

本年度、平成 22 年～23 年のデータについては新たに利用承諾がとれた医療機関のデータを加えて、分析用データセットの再作成を行うとともに、平成 24 年のデータについて分析用のデータセットの作成を行い、分析に供することができた。その際、キー情報の重複や必要なデータの欠損のある症例などは、分析に影響度が大きいので、それらの症例情報については確実に除去を行った。

このデータセットの完成で、DPC データの精度を向上させ、より高度な分析を実施することが可能となると考えられる。

A. 目的

本研究班において、収集した DPC データは、まず基本調査として入院情報の D/E/F ファイル・様式 1・3・4 ファイル（以下それぞれ FF1/FF3/FF4 ファイルと表記）と多岐にわたっている。また各病院のデータ収集が通年化され収集時期の範囲も広く、多数の医療機関が参加している。さらに、付随する調査として ICU/ER・リハビリ・CP・外来機能といった各種調査が付随し、それぞれが、基本調査の集計データを分析にあたって必要としている。

平成 25 年 3 月現在、本研究班では平成 15 年度調査から平成 21 年度調査までの 7 年度の調査データを医療機関との契約の元で保有し、分析に活用しているが、平成 24 年調査データの分量はすべてテキストのデータで、1TB を超えた量となっており、一般的な研究者が保有する分析環境（コンピュータの能力やデータを保管するストレージの量等）では処理が行えない状況となっている。

そこで、本研究では、平成 21 年度より本研究班の保有する DPC 調査データから、