

総括研究報告書

ICT・AI技術を活用した医療情報・ビッグデータ (腎臓病データベース) 解析技術の開発と
医療の質向上への貢献

研究代表者 柏原直樹 川崎医科大学 教授

研究要旨

日本腎臓学会 (JSN) は日本医療情報学会 (JAMI) と共同し、厚生労働省臨床効果データベース事業として全国規模の包括的慢性腎臓病 (CKD) 臨床効果情報データベース (J-CKD-DB) の構築に着手した。CKD は eGFR 60mL/分/1.73 m²未満、あるいはタンパク尿 (+) で定義され、個々の腎疾患を包含する広範な概念である。全国に約 1300 万人の有病者が推計され、増加が危惧されている。我が国では厚生労働省標準として電子カルテのデータを標準形式で格納する SS-MIX2 標準化ストレージがある。J-CKD-DB は SS-MIX2 標準化ストレージから CKD 該当例のデータ (患者基本情報、処方、臨床検査、等) を自動抽出してデータベース (DB) 化するものである。2019 年 3 月末時点で 15 施設・148,173 人の DB 構築に成功した。一方、JSN は腎臓病に関する規模の異なる複数の DB を構築してきた。(1)腎生レジストリ (Japan Renal Biopsy Registry : J-RBR) ; UMIN を介して手入力で腎生検症例を登録、(2) 各種腎疾患 DB (J-RBR から生成) 等の疾患単位 DB である。(1)、(2)はいずれも Web を用いた手作業での入力であるため入力負荷が大きく、数十万人規模以上の DB 構築が困難等の課題に直面していた。

本研究では、上記課題を克服し、腎臓病に関する全国規模の包括的データベースを構築し、腎臓病の実態調査、予後規定因子の解析、腎臓病診療の質向上、健康寿命延伸に寄与することを目的とした。

今後、本DBを活用して以下に取り組む。

- 1) 全国規模のCKD患者を対象にしたコホートを作成することで、本邦のCKD診療の実態調査、横断・縦断研究等を実施し、死亡、心血管系疾患、末期腎不全等のアウトカムの発症頻度およびそれらのリスク因子を評価する。費用対効果分析を行うために必要な基礎資料を作成する。
- 2) 研究課題を公募し、多くの1, 2次研究を支援する。
- 3) 得られたエビデンスをガイドライン改定に活かし、医療の質向上に貢献する。ガイドラインの普及、遵守率、阻害因子を分析する。アウトカムだけではなく、医療のプロセス、ガイドラインが推奨する標準医療への準拠率をQuality Indicator(QI)を設定して測定することで、医療の質評価が可能となる。
- 4) 以上の過程を円環的に循環させることで、遵守可能で完成度の高いガイドラインの作成が可能となり、診療の質向上に貢献可能とする。
- 5) マルコフモデルを用いた医療経済的アプローチが可能となる。腎臓病の病態推移確率を容易に、かつより精度高く算出可能となり、医療経済分析をより高い精度をもって進める。
- 6) 将来的に医療等IDが導入され、ナショナルレセプトデータベースへの連結、レセプト情報との突合が可能となれば、医療費負担の実態、ステージ毎の医療費負担の状況が把握でき、医療行為、薬剤療法等に関する正確な費用対効果分析が可能となる。

DB構築課程で、多施設EHRを用いたデータベース構築に際しての複数の課題に遭遇し、解決法を構築した。このノウハウは経験知として他データベース構築に際しても活用可能である。

本研究は、EHRの利活用方法の構築の端緒となるものであり、EHR活用の好事例となりえるものである。

研究分担者

南学正臣 東京大学 教授

大江和彦 東京大学 教授

岡田美保子 一般社団法人医療データ活用基盤整備機構 代表理事

片岡浩巳 川崎医療福祉大学 教授

神田英一郎 川崎医科大学 特任教授

A. 研究目的

腎疾患患者は年々増加傾向にあり、加齢とともに腎機能は低下するため、超高齢化社会である本邦では、一層の増加が危惧される。平成 30 年末には約 33 万人が透析療法を受けているが、心不全等の各種合併症のため、生命予後は必ずしも良好ではない(死因の第 7 位)。透析医療費は総額約 1.8 兆円(平成 26 年)にのぼり、医療財政の圧迫要因ともなっている。腎臓病の有効な予防、治療法開発、正確でより精緻な予後予測法の確立は喫緊の課題である。

また透析の高リスク者である慢性腎臓病(Chronic Kidney Disease:CKD)患者は約 1200 万人に上ると推計されている。CKD は eGFR 60mL/分/1.73 m²未満、あるいはタンパク尿(+)で定義され、個々の腎疾患を包含する概念である。厚生労働省健康局では、「腎疾患対策検討会(座長: 柏原直樹)」を立ち上げ、2018 年 7 月報告書を発出した。そこでは、今後 10 年間で新規導入患者数を 10%低減させることが目標として掲げられている。CKD は循環器疾患、糖尿病と同様に、WHO が健康寿命を損なう重大要因である Non-Communicable Disease (NCD) としてグローバルに注力する疾患である。

日本腎臓学会は日本医療情報学会と共同し、厚生労働省臨床効果データベース事業に採択(26 年)され、全国規模の包括的 CKD 臨床効果情報データベース(J-CKD-DB)構築に着手し、28 年から 3 年間、同臨床研究等 ICT 基盤構築研究事業にも採択され、データベースを構築した。SS-Mix2 を活用し、電子カルテ情報から CKD 該当例を検出するアルゴリズムを構築し、網羅的に患者データ(患者基本情報、処方、検査値等)を自動抽出し大規模な DB を構築した。

31 年 3 月末まで全国 15 病院、14 万 3 千人以上からなる大規模な CKD データベースを完成するに至った。現在、従来型の統計手法を用いて、解析(年齢・性別・重症度別の有病率等)を行っている。

本研究では、上記課題を克服し、腎臓病に関する全国規模の包括的データベース J-CKD-DB およびそれに後続する縦断データベース(J-CKD-DB-Ex)を用いて、腎臓病の実態調査、予後規定因子の解析、腎臓病診療の質向上、健康寿命延伸に寄与することを目的とする。

B. 研究方法

本研究事業では J-CKD-DB および J-CKD-DB-Ex を使用し解析を行う。

1. J-CKD-DB (-Ex) のデータ項目

表 1 に J-CKD-DB で収集しているデータ項目(一部省略)を示す。これらはすべて SS-MIX2 標準

化ストレージに保存される項目である。SS-MIX2 標準化ストレージは、データを格納するための仕様とともに、病院情報システムにおけるメッセージ(オーダ)の形式として HL7 V2.5 を指定し、医薬品については HOT コード、臨床検査については JLAC10 コードを標準としている。

SS-MIX2 には標準化ストレージと拡張ストレージがあるが、本研究では拡張ストレージは用いていない。以下では、SS-MIX2 標準化ストレージを、単に SS-MIX2 と表す。

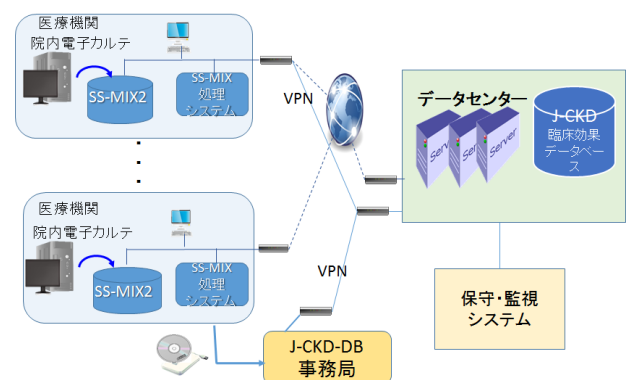
2. データベース設計・管理

本研究では東京大学で開発された多目的臨床データ登録システム(Multi-purpose Clinical Data Repository System: MCDRS)をデータベース構築に用いている。MCDRS は Web ベースの臨床症例データ登録システム用のソフトウェアである。MCDRS は、SS-MIX2 との連携機能を有し、SS-MIX2 からのデータ収集も想定されている。

3. データ抽出・収集の方法

データは、SS-MIX2 標準化ストレージからプログラムを用いて自動抽出する。このため、参加施設は SS-MIX2 が導入されていることが前提となる。図 1 に J-CKD-DB 登録システム全体の概要を図に示す。

図 1 J-CKD-DB システム全体概要



抽出基準と期間は次のとおりである。

・抽出基準：

eGFR 60mL/分/1.73 m²未満、またはタンパク尿(+)

・対象期間：

J-CKD-DB: 2014 年 1 月 1 日～12 月 31 日

J-CKD-DB-Ex: 2014 年 1 月 1 日～

SS-MIX2 から抽出して出力する時点で匿名化処理を行う。この一連の処理は、すべて施設内にて行われる。参加施設にて可搬媒体に出力し、可搬媒体を J-CKD-DB 事務局に送付して、事務局からデータベース登録を行う。

表 1 J-CKD-DB で収集するデータ項目

	項目 [単位等]
1	生年月
2	性別
3	受診科
4	例外登録・ 特殊登録 血液透析症例、腹膜透析症例、腎移植 症例、腎生検、J-RBR [1:該当、0:非該当]
5	治療開始日
6	患者区分
7	転帰区分
8	入院日時
9	退院日時
10	検査実施日時
11	血清クレアチニン [mg/dl]
12	尿蛋白 (定性) [-、±、+、2+、3+以上]
13	尿潜血 (定性) [-、±、+、2+、3+以上]
14	尿蛋白/クレアチニン比 [g/gCre]
15	尿蛋白(1日量) [g/日]
16	尿蛋白 (定量) [mg/dl]
17	尿アルブミン/クレアチニン比 (随時尿) [mg/gCre]
18	尿アルブミン (1日量) [mg/日]
19	尿クレアチニン (随時尿) [mg/dl]
20	尿クレアチニン (蓄尿) [mg/dl]
21	尿ナトリウム (随時尿) [mEq/l]
22	尿ナトリウム (蓄尿) [mEq/l]
23	尿・尿素窒素 (随時尿) [mg/dl]
24	尿・尿素窒素 (蓄尿) [mg/dl]
25	尿量 [ml/日]
26	血清総蛋白 [g/dl]
27	血清アルブミン [g/dl]
28	尿酸 [mg/dl]
29	尿素窒素 [mg/dl]
30	血清ナトリウム [mEq/l]
31	血清カリウム [mEq/l]
32	血清クロール [mEq/l]
33	血清マグネシウム [mg/dl]
34	総コレステロール [mg/dl]
35	HDL コレステロール [mg/dl]
36	LDL コレステロール [mg/dl]
37	中性脂肪 [mg/dl]
38	HbA1c [% (NGSP)]
39	グリコアルブミン [%]
40	シスタチン C [mg/l]
・	・
・	・ (省略)
・	・
67	血清補体価 (CH50) [U/ml]

68	内服薬: 処方日(オーダー)、医薬品名 (HOT コード)、 投与量、投与経路、投与期間
69	注射薬: 処方日 (実施)、医薬品名 (HOT コード)、 投与量、投与経路
70	外用薬: 処方日 (オーダー)、医薬品名 (HOT コード)、 投与量、投与経路
71	病名 [ICD10]

C. 研究結果

1. J-CKD-DB および J-CKD-DB-Ex の登録状況

J-CKD-DB は 21 施設からの参加を得て、データベース構築を進めている。平成 30 年度は 15 施設からの登録(もしくは再登録)があり、総登録件数は 148,173 件となっている。

J-CKD-DB-Ex へ 8 施設で倫理審査が受理されており、データベース構築を進めている。川崎医科大学を含む 2 施設で 2014 年から 2018 年末までの縦断データを抽出している。

2. J-CKD-DB 構築における問題と克服

開発を通じて見出された課題について以下に纏める。

(1) SS-MIX2 標準化ストレージの整備

SS-MIX2 は大学病院はじめ、大病院を中心に導入されている。導入目的、導入時期により、あるいは施設により、SS-MIX2 への出力状況は異なっている面がある。例えば標準化ストレージへの出力が継続的に行われていない、などの問題が挙げられる。

(2) 医薬品標準コード

臨床データベースでは医薬品情報は必須であり、SS-MIX2 では医薬品識別のため HOT コードを使用している。HOT コードは、厚生労働省標準となっているが、各施設では病院情報システム導入時から薬剤マスターを有しており、特段の必要性がなければ HOT コードは使われることはないため、コードの対応付けを行って来た。また、薬剤の同定のみならず、薬効分類が必要であり、重要な薬剤に関しては目視による対応付けも実施した。例を挙げると、降圧薬の有無を検索したいと考えた際に降圧薬に分類される薬剤の分類を行う必要がある。多くの薬剤は容易に分類できるが一部の利尿薬(サイアザイド系利尿薬やアルドステロン受容体拮抗薬など)は降圧薬と分類をし直すかどうか勘案する必要がある。これらの薬効分類を糖尿病治療薬、降圧薬、利尿薬、りん吸着薬など腎疾患研究に必要性が高い薬剤の整備を行うことにより運用可能となっている。

(3) 臨床検査の標準コード

SS-MIX2 では臨床検査コードとして、JLAC10 を推奨している。JLAC10 は日本臨床検査医学会が制定するコードで、次の 5 つの要素からなる 17

桁のコードである：1)分析物(5桁)、2)識別(4桁)、3)材料(3桁)、4)測定法(3桁)、5)結果識別(2桁)。(識別コードは分析物コードを、検査内容に沿って細分化する必要がある場合に、コードを付したものの。)

各施設は臨床検査のマスターを有しており、オーダには院内コード(ローカルコード)が利用される。JLAC10は病院のマスターには、ほとんど採用されていないため、ローカルコードへのJLAC10コードの割当てが必要となった。J-CKD-DBで収集する検査項目は50項目程度であり、本研究では、あらかじめJLAC10を割当てた表(1つの検査項目に複数のJLAC10コードが対応)を作成し、各施設にローカルコードへのJLAC10の割り当てを依頼した。しかし、臨床検査部門においても、JLAC10のコード割当ては時間を要する作業となっている。JLAC10コードは必ずしもユニークに決まらない場合もあり、課題として残っている。

(4) データクレンジング

臨床データベース構築においては、データのクレンジングが必須である。データクレンジングの考え方は、分野を問わず共通であるが、ここではJ-CKD-DBにおける主な点を述べる。

1) 検査値の単位

表1に示すとおり、検査項目には単位を定めており、参加施設からは検査値とともに単位を収集している。施設により、表1の単位と異なる場合があり、その場合はデータ変換を行っている。

また、単位が得られていない場合があり、参加施設より確認を得ている。

2) ローカルコードと標準コード

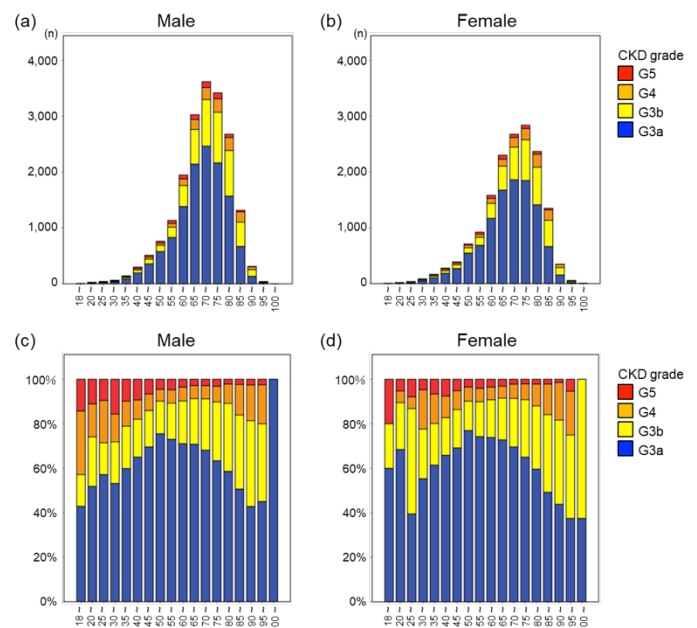
前述のとおり、病院では検査や薬剤のマスターにローカルコードが使われているが、J-CKD-DBではSS-MIX2の標準に準拠して、臨床検査にはJLAC10コード、薬剤にはHOTコードを用いることとしている。すでにマスターが標準コードに対応している一部の施設を除き、ほとんどの場合はローカルコードと標準コードとの対応付け作業が発生した。

SS-MIX2への出力ではローカルコードと標準コードをともに出力している。抽出には標準コードを抽出して、データベース登録することとしているが、標準コードではなく、ローカルコードが抽出されている場合があり、変換作業が発生している。また、医薬品に関しては、HOT9を用いることとしたが、HOT7、HOT9、HOT13なども見られた。これについては、分析時に、桁を揃えることになる。

3. J-CKD-DB 解析による診療実態の解明

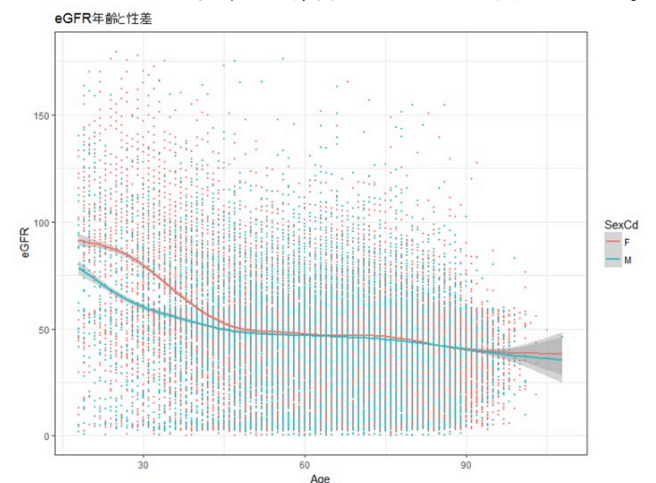
J-CKD-DBを元に本邦のCKDの実態を検討した。下図は男女別に年齢層毎のCKD各ステージの有病率を示したものである。

男性では60歳以上で、女性では65歳以上になるとG3b以降のCKD患者が増加することが示されている。

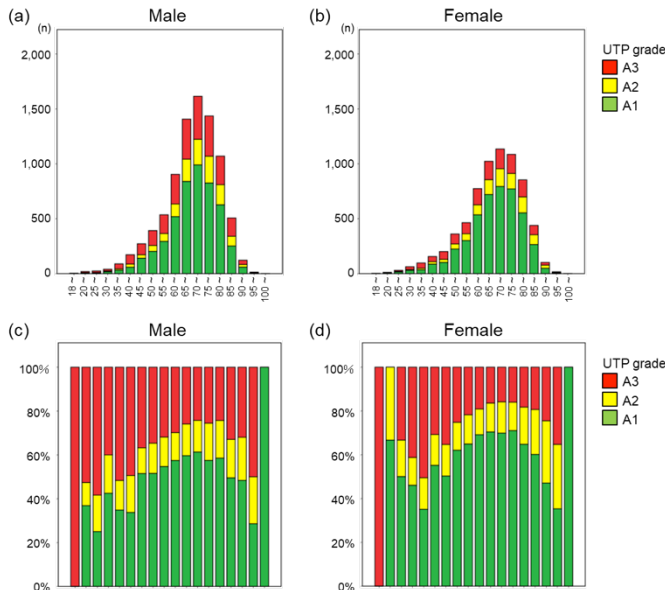


また加齢により腎機能が低下することが判明している。本邦は超高齢化社会を迎えているが、高齢化が腎機能に与える影響の詳細は不明である。

J-CKD-DBを用いて加齢に伴う腎機能低下への影響を検討したところ、男女別、年齢毎の推算GFR(eGFR)の変化を解析したものが下図である。



さらに、アルブミン尿、蛋白尿の年齢層別の解析を実施した。GFR との異なり若年層に蛋白尿の多い群が存在することが判明した。CKD 重症度分類の各層の有病率を解析した(下図)。



年齢層毎に重症度を分類し、各クラスでの有病率を解析したものが下図である。

Male Total					Female Total				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	3552 (41.2%)	762 (8.8%)	887 (10.2%)	5196 (60.3%)	G3a	3100 (45.5%)	540 (7.9%)	440 (6.5%)	4080 (59.8%)
G3b	1094 (12.7%)	403 (4.7%)	785 (9.1%)	2282 (26.5%)	G3b	1039 (15.2%)	290 (4.3%)	397 (5.8%)	1726 (25.3%)
G4	209 (2.4%)	124 (1.4%)	548 (6.4%)	881 (10.2%)	G4	324 (4.6%)	133 (1.9%)	326 (4.8%)	783 (11.5%)
G5	13 (0.2%)	11 (0.2%)	284 (3.3%)	265 (3.1%)	G5	26 (0.4%)	25 (0.4%)	178 (2.6%)	229 (3.4%)
Subtotal	4868 (56.4%)	1307 (15.2%)	2449 (28.4%)	3216/8624 (37.3%/100%)	Subtotal	4489 (65.8%)	988 (14.5%)	1341 (19.7%)	2139/6818 (31.4%/100%)

Male 18-44y					Female 18-44y				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	98 (12.2%)	28 (8.9%)	77 (10.7%)	198 (56.9%)	G3a	113 (31.3%)	28 (7.8%)	38 (11.7%)	205 (59.8%)
G3b	16 (4.6%)	17 (4.9%)	42 (12.4%)	76 (21.8%)	G3b	39 (10.8%)	15 (4.2%)	31 (8.6%)	85 (23.5%)
G4	4 (1.1%)	8 (2.3%)	39 (11.2%)	51 (14.7%)	G4	20 (5.5%)	6 (1.7%)	26 (7.2%)	52 (14.4%)
G5	1 (0.3%)	1 (0.3%)	21 (6.0%)	23 (6.6%)	G5	0 (0.0%)	4 (1.1%)	15 (4.2%)	19 (5.3%)
Subtotal	119 (34.2%)	54 (15.5%)	175 (50.3%)	206/348 (59.2%/100%)	Subtotal	172 (47.6%)	53 (14.7%)	136 (37.7%)	361/561 (64.3%/100%)

Male 45-64y					Female 45-64y				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	920 (43.7%)	187 (8.9%)	277 (13.2%)	1384 (65.8%)	G3a	909 (50.5%)	137 (7.6%)	154 (8.6%)	1200 (66.7%)
G3b	201 (9.6%)	65 (3.1%)	192 (9.1%)	458 (21.8%)	G3b	189 (10.5%)	60 (3.3%)	114 (6.3%)	363 (20.2%)
G4	30 (1.4%)	15 (0.7%)	137 (6.5%)	182 (8.7%)	G4	58 (3.2%)	23 (1.3%)	95 (5.3%)	176 (9.8%)
G5	3 (0.1%)	6 (0.3%)	71 (3.4%)	80 (3.8%)	G5	5 (0.3%)	8 (0.4%)	48 (2.7%)	61 (3.4%)
Subtotal	1154 (64.5%)	273 (13.0%)	677 (32.2%)	2067/2104 (98.3%/100%)	Subtotal	1161 (64.5%)	228 (12.7%)	411 (22.8%)	1800 (82.8%/100%)

Male 65-74y					Female 65-74y				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	1390 (46.0%)	260 (8.6%)	272 (9.0%)	1922 (63.6%)	G3a	1095 (50.7%)	172 (8.0%)	97 (4.5%)	1364 (63.2%)
G3b	371 (12.3%)	137 (4.5%)	268 (8.9%)	776 (25.7%)	G3b	330 (15.3%)	78 (3.6%)	108 (5.0%)	516 (23.9%)
G4	64 (2.1%)	34 (1.1%)	156 (5.2%)	254 (8.4%)	G4	82 (3.8%)	38 (1.8%)	86 (4.0%)	206 (9.5%)
G5	4 (0.1%)	5 (0.2%)	61 (2.0%)	70 (2.3%)	G5	6 (0.3%)	9 (0.4%)	57 (2.6%)	72 (3.3%)
Subtotal	1829 (60.5%)	436 (14.4%)	757 (25.0%)	3022 (33.1%/100%)	Subtotal	1513 (70.1%)	297 (13.8%)	348 (16.1%)	2158 (26.0%/100%)

(i) 75-84y					(j) 75-84y				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	978 (39.0%)	229 (9.1%)	207 (8.3%)	1414 (56.4%)	G3a	836 (43.1%)	150 (7.7%)	97 (5.0%)	1083 (59.8%)
G3b	398 (15.9%)	149 (5.9%)	210 (8.4%)	757 (30.2%)	G3b	364 (18.8%)	90 (4.6%)	102 (5.3%)	556 (28.7%)
G4	72 (2.9%)	45 (1.8%)	146 (5.8%)	263 (10.5%)	G4	114 (5.9%)	44 (2.3%)	85 (4.4%)	243 (12.5%)
G5	3 (0.1%)	5 (0.2%)	64 (2.6%)	72 (2.9%)	G5	10 (0.5%)	2 (0.1%)	46 (2.4%)	58 (3.0%)
Subtotal	1451 (57.9%)	428 (17.1%)	627 (25.0%)	2506 (36.0%/100%)	Subtotal	1324 (68.2%)	286 (14.7%)	330 (17.0%)	1940 (30.4%/100%)

(k) 85y-					(l) 85y-				
Stage	A1	A2	A3	Subtotal	Stage	A1	A2	A3	Subtotal
G3a	166 (25.8%)	58 (9.0%)	54 (8.4%)	278 (43.2%)	G3a	147 (26.3%)	53 (9.5%)	28 (5.0%)	228 (40.8%)
G3b	108 (16.8%)	35 (5.4%)	72 (11.2%)	215 (33.4%)	G3b	117 (20.9%)	47 (8.4%)	42 (7.5%)	206 (36.9%)
G4	39 (6.1%)	22 (3.4%)	70 (10.9%)	131 (20.3%)	G4	50 (8.9%)	22 (3.9%)	34 (6.1%)	106 (19.0%)
G5	2 (0.3%)	1 (0.2%)	17 (2.6%)	20 (3.1%)	G5	5 (0.9%)	2 (0.4%)	12 (2.1%)	19 (3.4%)
Subtotal	315 (48.9%)	116 (18.0%)	213 (33.1%)	644 (48.4%/100%)	Subtotal	319 (57.1%)	124 (22.2%)	116 (20.8%)	559 (43.2%/100%)

以上の結果をとりまとめ、発表した (Sci Rep. 2020 Apr 30;10(1):7351)。

4. 縦断データ (J-CKD-DB-Ex) 解析による予後予測アルゴリズムの開発と予測因子の同定

J-CKD-DB のデータベース構築システムを用い 2014年から2018年12月31日までの5年間のデータ取得を行なっている (J-CKD-DB-Ex)。

現在、J-CKD-DB-Ex として川崎医科大学含め8施設の倫理審査が終了している。先行して川崎医科大学における5年間のデータ解析を行なった。

①時系列データ解析による腎機能予後の予測:

J-CKD-DB-Ex では5年間の縦断データが正確に収集可能である。これを活用し時系列データの解析を Recurrent Neural Network:RNN を用いて行う。

通常のニューラルネットワークでは、ある層の出力は、次の層の入力に利用されるのみである。しかし RNN では、ある層の出力は、次の層の入力として利用されるだけでなく、一般的なニューラルネットワークの最後の層のような (中間データではないユーザーが利用可能な) 出力としても利用される。また、各層の入力として、前の層の入力のみではなく、時系列のデータポイントも入力とする。さらに直前の入力層のみならず、離れた地点での入力層の影響を加味することが可能となった。代表的なアルゴリズムとして Long short-term memory (LSTM) が挙げられる。

これらを用いることで、腎機能予後を予測することが可能となる。5年間の縦断データの一部を学習セットとして、残りを検証データとして用いてモデルを作成する予定である。

②腎機能予後規定因子の同定:

微量元素である亜鉛に着目し研究を行っている。低亜鉛血症は貧血の原因として認識されているが、一方で腎組織障害との関連が想定される。しかしながら低亜鉛血症が腎疾患進展にどのような影響を与えるかは不明である。傾向スコア (プロペンシティブスコア) を使用した解析で低亜鉛血症が腎代替療法導入および死亡のリスクであることが示さ

れた（論文投稿中）。今後は亜鉛薬投与の効果を実用 world data を用いて検討を行う。

③高カリウム血症のリスク同定

腎疾患診療においてカリウム値の管理は重要な課題である。カリウム値が一定以上となると致命的不整脈のため死に至る。また高カリウム血症は入院理由の一つでもある。カリウム値に影響を与える因子は多く、RAS 阻害薬などの治療薬や患者背景（高齢者など）などが指摘されており、正確に予測することは困難である。このため本データベースを使用し高カリウム血症の予測を、機械学習を用いて行う。

D. 考察

1. 電子カルテ情報（HER）を用いたデータベース構築の課題

本プロジェクトではSS-MIX2を用いて自動的に臨床データを収集するものである。データベース構築過程で様々な課題解決を行ってきた。

①. 薬剤情報の整備

各施設において HOT コードの対応付けを行っていることは、結果に述べたとおりである。

理論上、HOT コードでは一薬剤に一つのコードが振られるため、コードが特定されれば、単一薬剤を抽出することが可能である。

しかしながら研究使用においては、薬効単位（降圧薬、糖尿病薬等）で抽出することが必要な場合が多い。このため、薬効分類を行う必要がある。

例えばアムロジピンという一般名の薬剤には HOTA コードが複数振られている。これは用量の違いのみならず、販売ベンダーの違いなどによる。このためアムロジピンの使用患者を抽出する際には、全てのコードのリストが必要になる。さらに研究ではアムロジピンのみを対象として抽出することは稀で、降圧薬の薬効で抽出することが一般的である。これらのクエリに耐えうる薬効分類を作成する必要がある。これは手作業で担うところが大きく、時間を要した。

今後は、研究課題の拡大に合わせて、さらなる薬効分類の充実が必要と考えている。

②. 臨床検査データ

J-CKD-DB および J-CKD-DB-Ex の構築により臨床検査の値には、基準値の問題がある。専門団体や地域、あるいは国立大学病院による取り組み等があるが、いまだ全国で統一されておらず施設間で差があることがわかってきた。

SS-MIX2 では、臨床検査データの定性結果、不等号の表記について推奨があるが、必ずしも順守されていない場合がある。これらを補正しデータベースの精度を上げてきた。

しかしながら今後は新たな検査項目の追加や新規施設において調整が必要である。これらを根本的に解決するべく、二次活用を視野に置いた臨床検査データの表記について改善を急ぐべきであると考えている。

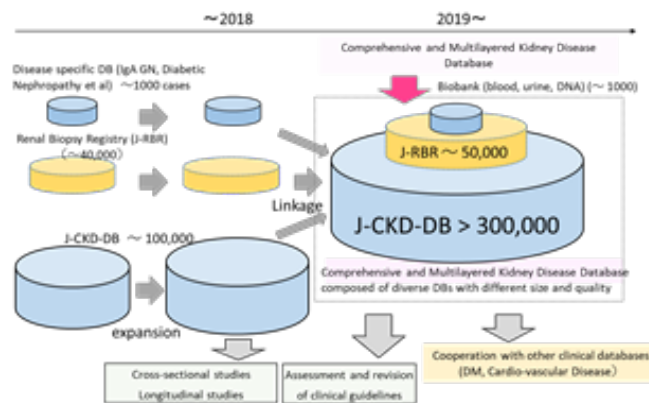
2. 院内情報システムからの取得が困難な情報

CKD は eGFR 60mL/分/1.73 m²未満あるいはタンパク尿 (+) で定義され、個々の腎疾患を包含する広範な概念である。J-CKD-DB の研究目的に照らして、次に該当する場合は DB 上でフラグを立てることとした。「1 血液透析症例 2 腹膜透析症例 3 腎移植症例 4 腎生検 5 J-RBR」。これらの情報は、SS-MIX2 の範囲外であり、さらには電子カルテシステム、レセプト等でも完全には把握できない。特にイベント発生に関しては十分ではない。海外に比較し、本邦では病名の信頼性が低く疾患把握が容易ではない。これらを解決するべく今後はバリデーション研究を推進し、データベース研究の更なる利活用の推進を図る必要がある。

E. 今後の展望

J-CKD-DB (-Ex) をさらに拡充すると同時に本事業での研究活用を介して、整備を行う。

他の DB との連結を強化する。最終的には、ゲノム、血液・尿等の生体サンプルを連結した重層的な DB を構築し、今後の研究に資することを旨とする。



この重層的 DB を活用して以下に取り組む。

- 1) 全国規模の CKD 患者を対象にしたコホートを作成することで、本邦の CKD 診療の実態調査、横断・縦断研究等が可能になる。死亡、心血管系疾患、末期腎不全等のアウトカムの発症頻度およびそれらのリスク因子（重症化の要因分析）を評価可能である。費用対効果分析を行うために必要な基礎資料を作成する。
- 2) 厚生労働科学研究難治性疾患等政策研究事業である「難治性腎疾患に関する調査研究」の対象疾患である IgA 腎症、急速進行性糸球体腎炎、ネフローゼ症候群、多発性嚢胞腎の予後調査等の前向き研究を支援する。

- 3) 研究課題を公募する事で多くの1, 2次研究が実施可能であり、本邦からのエビデンス創出を加速するエンジンとなる。
- 4) 得られたエビデンスをガイドライン改定に活かし、医療の質向上に貢献できる。
- 5) ガイドラインの普及、遵守率、阻害因子を分析することができる。
- 6) アウトカムだけではなく、医療のプロセス、ガイドラインが推奨する標準医療への準拠率をQuality Indicator(QI)を設定して測定することで、医療の質評価が可能となる。
- 7) Evidence-Practice gap,地域による医療の質のばらつきも評価可能となる。
- 8) 以上の過程を円環的に循環させることで、遵守可能で完成度の高いガイドラインの作成が可能となり、診療の質向上に貢献可能である。
- 9) マルコフモデルを用いた医療経済的アプローチが可能となる。腎臓病の病態推移確率を容易に、かつより精度高く算出可能となり、医療経済分析をより高い精度をもって進めることが可能となる。
- 10) 将来的に医療等 ID が導入され、ナショナルレセプトデータベースへの連結、レセプト情報との突合が可能となれば、医療費負担の実態、ステージ毎の医療費負担の状況が把握でき、医療行為、薬剤療法等に関する正確な費用対効果分析が可能となる。

F. 結論

HER を活用し 15 施設、148,183 件からなる大規模な全国規模の CKD データベース (J-CKD-DB) の構築に成功した。さらに 5 年間の縦断データベースにも着手している (J-CKD-DB-Ex)。本データベースの活用により多くの研究が可能となり、今後の発展性は極めて大きい。当該領域の診療の質向上・均霑化に貢献可能である。

DB 構築課程で、多施設 EHR を用いたデータベース構築に際しての複数の課題に遭遇し、解決法を構築しえた。このノウハウは経験知として他データベース構築に際しても活用可能である。

本研究は、EHR の利活用方法の構築の端緒となるものであり、EHR 活用の好事例となりえる。

G. 研究発表

1. 論文発表

- 1) Naoki Nakagawa, Tadashi Sofue, Eiichiro Kanda, Hajime Nagasu, Kunihiro Matsushita, Masaomi Nangaku, Shoichi Maruyama, Takashi Wada, Yoshio Terada, Kunihiro Yamagata, Ichiei Narita, Motoko Yanagita, Hitoshi Sugiyama, Takashi Shigematsu, Takafumi Ito, Kouichi Tamura, Yoshitaka

Isaka, Hirokazu Okada, Kazuhiko Tsuruya, Hitoshi Yokoyama, Naoki Nakashima, Hiromi Kataoka, Kazuhiko Ohe, Mihoko Okada, Naoki Kashihara. J-CKD-DB: A Nationwide Multicentre Electronic Health Record-Based Chronic Kidney Disease Database in Japan. *Sci Rep* 10(1):7351.2020

- 2) Kanda E, Nangaku M. Are SGLT2 inhibitors a targeted treatment for diabetic kidney disease? *Kidney Int.* 2019 Jul;96(1):8-10.
- 3) Fukui A, Yokoo T, Nangaku M, **Kashihara N**. New measures against chronic kidney diseases in Japan since 2018. **Clin Exp Nephrol.** 23(11):1263-1271,2019
- 4) 長洲一、柏原直樹：腎臓病診療におけるデータベース構築(J-CKD-DB)の現状. *Pharma Medica* 38(1):71-75,2020
- 5) 柏原直樹：高血圧診療におけるAI/ビッグデータ/モバイルテレメディシンの活用. *血圧* 26(12):29-33, 2019
- 6) 柏原直樹：30万人規模を目指す 国内CKDデータベースが中間解析(学会報告記事). *Medical Tribune* 52(16):8-8,2019
- 7) 柏原直樹、長洲一、神田英一郎。糖尿病合併症とビッグデータ構築. *Precision Medicine* 2(5):402-405,2019
- 8) 岡田美保子.耳鼻咽喉科等におけるビッグデータ、リアルワールドデータ活用の動向.*JOHN S* 35(10):1433-1436,2019

2. 学会発表

- 1) 祖父江理、中川直樹、神田英一郎、長洲一、岡田浩一、岡田美保子、片岡浩巳、柏原直樹。外来 CKD 患者における高尿酸血症・電解質異常の有病率・QI 遵守率に関する解析：J-CKD データベース研究一次解析.第 62 回日本腎臓学会学術総会 2019 年 6 月 23 日 名古屋
- 2) 中川直樹、祖父江理、神田英一郎、長洲一、岡田浩一、岡田美保子、片岡浩巳、柏原直樹。外来 CKD 患者の実態調査：わが国における慢性腎臓病データベース(J-CKD-DB)一次解析. 第 62 回日本腎臓学会学術総会 2019 年 6 月 23 日 名古屋
- 3) 祖父江理、中川直樹、神田英一郎、長洲一、岡田浩一、岡田美保子、片岡浩巳、柏原直樹、J-CKD-DB 委員会。外来 CKD 患者における貧血有病率、・QI 遵守率に関する解析：J-CKD データベース研究一次解析.第 62 回日本腎臓学会学術総会 2019 年 6 月 22 日 名古屋
- 4) 柏原直樹.J-CKD-DB を CKD 診療ガイドライン診療に活かす. 第 62 回日本腎臓学会学術総会 2019 年 6 月 22 日 名古屋
- 5) 岡田美保子. ビッグデータは医療をどのよう

に変わるか? その先にある医療の姿を考える. 日本医療・病院管理学会第 377 回例会(東海病院管理学会併設)抄録, 日本医療・病院管理学会誌, 56(4), 192, 2019.

- 6) 岡田美保子. 健康・医療リアルワールドデータの活用-課題解決と成果の普及にむけて-, 第 10 回横幹連合コンファレンス, 健康医療における臨床試験データからリアルワールドデータへ ~ データ駆動型ヘルスデータサイエンスの新しい潮流~, 長岡, 2019 年 11 月 30 日.
- 7) 土屋文人, 古川裕之, 副島秀久, 大箸義章, 岡田美保子. RWD 時代に医療情報が超えなければならない壁を考える-薬剤に関する情報を中心に-. 医療情報学 39(Suppl.), 133-134, 2019.
- 8) 岡田美保子. リアルワールドデータ活用の時代にこそ求められる医療 CIO. 医療情報学 39(Suppl.), 47-48, 2019.
- 9) 岡田美保子. 大規模医療データ活用の潮流 -リアルワールドデータとモバイルヘルスを中心に. 第 1 回 Digital Hypertension Conference, 「ビッグデータの活用で変わる高血圧診療の未来」, 東京, 2019 年 10 月 25 日.
- 10) 岡田美保子. いま問われる診療データの質

-医療情報技師への期待. 第 6 回日本医療情報学会東北支部会医療情報研修会, 秋田, 2019 年 7 月 14 日.

- 1 1) Mihoko Okada. New Age of Clinical Information Usage for Healthcare. Keynote Presentation, 2019 CDISC Japan Interchange, Tokyo, July 10, 2019.
- 1 2) 岡田美保子. 医療ビッグデータの利活用は医療にどう影響を及ぼすのか. 第 30 回日本医学会総会 2019 中部, 「1-1-1 ビッグデータは医療の何をどう変えるのか」, 名古屋, 2019 年 4 月 27 日.

H. 知的所有権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし