

平成29年度厚生労働科学研究費補助金（臨床研究等ICT基盤構築・人工知能実装研究事業）
総括研究報告書

腎臓病データベースの拡充・連携強化と包括的データベースの構築

研究代表者 柏原 直樹 川崎医科大学医学部腎臓・高血圧内科学 教授

研究要旨

日本腎臓学会は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100万人の有病者が推計され、増加が危惧されている。本研究では、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する。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該当例のデータを自動抽出してデータベース化するものである。平成29年度は、CKD該当例を自動抽出しうるアルゴリズムを作成し、電子カルテ情報(EHR: Electronic Health Record)から、SS-MIX2を活用し、多施設からCKD臨床情報を収集しDBを構築し、12病院から105,068人のCKD-DBを構築した。得られた検査値等のデータのクレンジング・バリデーション、標準化作業を行なった。また、SS-MIX2を用いて臨床データベースを構築する際の課題を抽出し、解決方法の開発に着手した。この成果は他研究にも活用できるものとなる。さらに、本邦のCKDの実態についての予備的解析を行った。J-CKD-DBの医学研究、疫学研究における意義、発展性は極めて大きく、遵守可能で完成度の高いガイドラインの作成を可能とし、診療の質向上・均霑化に貢献できると考える。

研究分担者

横山 仁・金沢医科大学腎臓内科学 教授
南学正臣・東京大学医学部附属病院腎臓・内分泌内科 教授
山縣邦弘・筑波大学医学医療系腎臓内科学 教授
和田隆志・金沢大学医薬保健研究域医学系腎臓内科学・腎臓病態統御学 教授
中島直樹・九州大学病院メディカル・インフォメーションセンター 教授
杉山 斉・岡山大学医歯薬学総合研究科血液浄化療法人材育成システム開発学 教授
丸山彰一・名古屋大学・大学院医学系研究科病態内科学講座腎臓内科学 教授
岡田美保子・公益財団法人神戸医療産業都市推進機構・医療イノベーション推進センター・客員研究員

A. 研究目的

本研究の目的は、ICT技術を活用して、CKDの実態調査、予後規定因子の解析、標準治療の均霑化率・準拠率の評価、有効な予防・重症化抑制策の立案、腎臓病診療の質向上、腎臓病の予後改善、透析導入率の減少、健康寿命延伸に寄与することである。

B. 研究方法

1. J-CKD-DB のデータ項目

J-CKD-DBで収集しているデータ項目(患者基本情報、処方、臨床検査、等)は、すべて厚生労働省標準である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を定めるとともに、病院情報システムにおけるメッセージ(オーダ)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 とも表す。

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)をデータベース設計に用いている。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り、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

参加施設はSS-MIX2が導入されていることを前提としている。J-CKD-DB に登録するデータはSS-MIX2 からプログラムを用いて自動抽出する。図1にJ-CKD-DB登録システム全体の概要を示す。

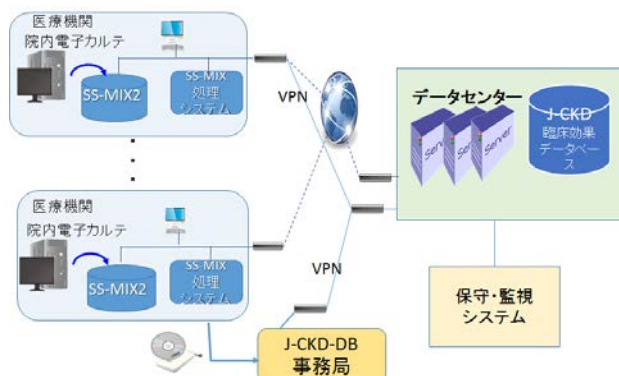


図1 J-CKD-DB システム全体概要

4. 抽出基準と対象期間

抽出基準および収集の対象とする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：

2014年1月1日～12月31日までの間

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。匿名化したデータはVPN接続により送

信するか、または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。

(倫理面への配慮)

本研究の実施にあたり、すべての研究者は個人情報保護法、ヘルシンキ宣言（2013年10月改正）、人を対象とする医学系研究に関する倫理指針（平成27年4月1日施行）、改正GCP省令（平成24年12月改正）、医療情報システムの安全管理に関するガイドライン第4.2版（2013年10月改正）等を厳格に遵守している。J-CKD-DBへの登録は匿名化した上でを行っている。インフォームドコンセントに関しては研究内容、研究に用いられる情報の利用目的についてHP等で公示し、該当する患者が拒否できる機会を保障してオプトアウト方式で実施している。J-RBRに関しては従来通り、本研究に関する説明文書を用意し、十分な説明を行った上で本研究への参加について同意を文書で得ている。個人情報の厳重な管理を行うため日本腎臓学会倫理委員会（委員長香美祥二）の下に「個人情報保護委員会」が設置されている。

C. 研究結果

1. データベース登録状況

J-CKD-DBではS-MIX2標準化ストレージ(以下SS-MIX2)を用いて、該当例のデータ(患者基本情報、処方、臨床検査、等)を抽出してデータベース登録している。表1に登録対象と、2018年3月現在の登録施設数、登録件数(症例数)を示す。

表1 登録対象*と登録状況

登録対象	対象期間	2014年1月1日～2014年12月31日までの間
	年齢	18歳以上
	登録基準	eGFR 60mL/分/1.73 m ² 未満、またはタンパク尿(1+以上)
登録状況	登録施設数	12施設
	登録施設	川崎医科大学、九州大学、香川大学、高知大学、和歌山県立医科大学、東京大学、新潟大学、金沢大学、旭川医科大学、島根大学、岡山大学、筑波大学
	総登録件数	105,068件(2018年3月現在)

*対象期間に一度でも基準に合致した場合に、2014年1月1日～2014年12月31日までのすべての登録項目のデータを登録している。

2. J-CKD-DB 構築・運用

J-CKD-DB の構築のためのデータ抽出、データ登録、データクレンジングの手順は以下のとおりである。図 2 に示す①～⑦の順に述べる。

データ抽出

- ①施設内のマスターに HOT コード、JLAC10 コードの対応付けを行う。
- ②施設内にて SS-MIX2 への再出力を行う。
- ③SS-MIX2 から登録データを匿名化して抽出する。抽出プログラムは参加施設に提供し、抽出処理は施設内で行う。

データ登録

- ④抽出されたデータは暗号化して媒体に保存し J-CKD-DB 事務局に送付する。事務局は MCDRS(クラウド版)を用いてアップロードする。データベース登録時にデータは復号される。事務局にて登録後のデータをエクスポートして確認する。

データクレンジング

- ⑤エクスポートしたデータに基本的なデータクレンジング処理を施す。
- ⑥検査データについて分布の視覚化、データ変換等の処理を行う(詳細については後述)。

データ活用

研究活用のための手順を定め、2018年3月より運用を開始した。データ活用については後述する。

3. 見出された課題と対処方法

2017 年度は登録データ全体に渡るクレンジングを実施し手順を確立した。

①院内における SS-MIX2

SS-MIX2 の整備状況の問題は昨年度報告しているが、まったく出力されていない項目を精査し、原因を特定した。例えば尿蛋白(定性)は 3 施設で出力されていなかった。施設側と協議して SS-MIX2 への再出力もしくは施設内 DWH により対処した。

②医薬品と検査項目へのコード割当

HOT コード割当て、および JLAC10 コードの付与は、いずれの施設でも相当の時間と労力を要するだけでなく完全な確定ができない。一つの研究班、一つの団体の取組では解決できず、国全体に渡る根本的な方策が必須であると考えられた。

③SS-MIX2 からの抽出

登録基準に用いる尿蛋白(定性)は、各施設より定性値の表記を得て抽出処理を行ったが、例えば「+、+1、+2、+3」とされていても、実際には「±」、「(+1)」、「1+」、「(1+) 30」、さらに全角文字など様々な表記揺れがみられ、正確な抽出ができていない施設があった。当初は再抽出を行う必要があったが、正確な表記の把握を徹底することにより、回避可能となった。

④データエクスポート

登録データのエクスポート(ダウンロード)処理に相当の時間を要し(およそ 1 ギガで 1 時間)、機能向上が必要となっている。

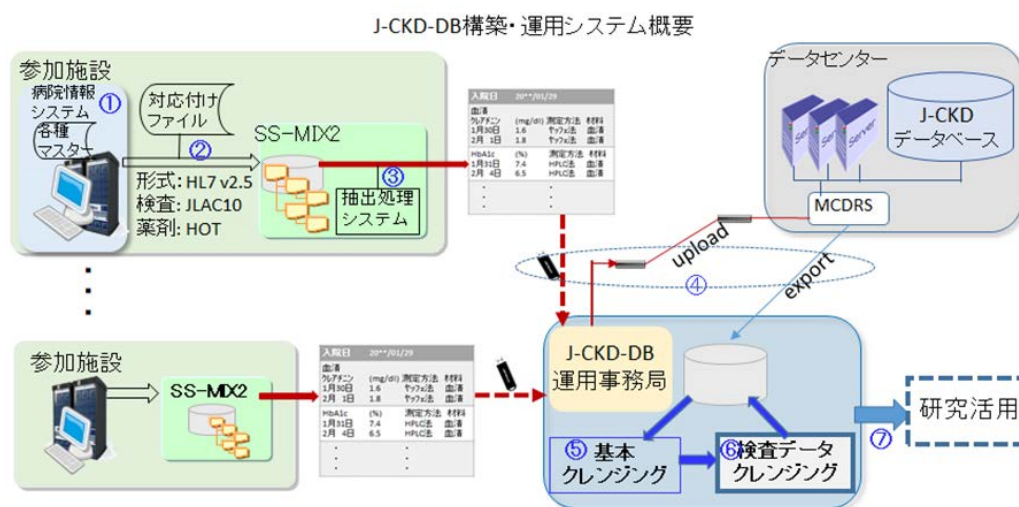


図 2 J-CKD-DB 構築・運用の概要

⑤基本クレンジング(薬剤)

全施設のデータにローカルコードが混在していた。1施設では全てローカルコードであった。HOT9を指定していたが、HOT7、HOT13も見られた。この問題に対しては、院内薬剤マスターの提供を受けて、HOTコードへの変換作業を行っている。HOTコードの桁が異なる場合は、分析時に桁を揃える。

⑤基本クレンジング(検査項目コード)

JLAC10コードとローカルコードが混在していた。一部に複数の検査項目に同じローカルコードが割当てられている事例もみられた。施設に問い合わせ、JLAC10の対応付けを行っている。

⑥検査については「4.検査データの精査」で述べる。

4.検査データの精査

J-CKD-DBは、登録項目に多くの尿検査の項目を含むことが特徴として挙げられる。また測定頻度が高くはない、免疫検査や内分泌検査などを含むことも特徴である。

4.1 検査データの単位

表2に登録項目のうち検査項目の一覧を示す。各施設で検査項目の単位が正しく出力されていた割合を表3に示す。単位が得られていない場合は施設より確認を得た。単位には、かなりのばらつきがあることがわかる。

電解質の単位として、mmol/Lが使用されている施設があった。血清・尿ナトリウムや血清カリウムであれば電荷が1であるためmEq/L、mmol/Lのいずれでも問題になることはないが、カルシウムの場合、電荷は2であるため値が実際と異なる例が生じる危険性がある。また、カルシウムでmg/dLを使用している施設があった。

表2 J-CKD-DB登録検査項目(その他の項目は省略)

項目	単位	項目	単位
尿蛋白(定性)		血清クレアチニン値	mg/dL
尿潜血(定性)		血清総蛋白	g/dL
尿蛋白/クレアチニン比	g/gCre	血清アルブミン	g/dL
尿蛋白(1日量)	g/日	尿酸	mg/dL
尿蛋白(定量)	mg/dL	尿素窒素	mg/dL
尿アルブミン/クレアチニン	mg/gCre	AST	U/L
尿アルブミン(1日量)	mg/日	ALT	U/L
尿クレアチニン(随時尿)	mg/dL	血清ナトリウム	mEq/L
尿クレアチニン(蓄尿)	mg/dL	血清カリウム	mEq/L
尿ナトリウム(随時尿)	mEq/L	血清クロール	mEq/L
尿ナトリウム(蓄尿)	mEq/L	血清カルシウム	mg/dL
尿・尿素窒素(随時尿)	mg/dL	血清リン	mg/dL
尿・尿素窒素(蓄尿)	mg/dL	血清マグネシウム	mg/dL
尿量	mL/日	総コレステロール	mg/dL
WBC	/μL	HDLコレステロール	mg/dL
RBC	/μL	LDLコレステロール	mg/dL
Hb	g/dL	中性脂肪	mg/dL
Ht	%	HbA1c	%
PLT	/μL	グリコアルブミン	%
CRP	mg/dL	血清鉄	μg/dL
抗核抗体-FA法		総鉄結合能(TIBC)	μg/dL
抗核抗体-EIA法		トランスフェリン飽和度	%
MPO-ANCA(EIA)		フェリチン	ng/mL
MPO-ANCA(CLEIA)	倍	Intact PTH	pg/mL
PR3-ANCA(EIA)		Whole PTH	pg/ml
PR3-ANCA(CLEIA)	U/mL	シスタチンC	mg/L
抗GBM抗体(EIA)	U/mL	重炭酸イオン濃度	mEq/L
抗GBM抗体(CLEIA)	U/mL	BNP	pg/mL
血清補体価(CH50)	U/mL		

表3 検査単位の一致割合

施設	出力項目数	単位一致項目数	割合
A	47	31	65.96%
B	46	15	32.61%
C	47	1	2.13%
D	58	10	17.24%
E	46	7	15.22%
F	41	4	9.76%
G	41	1	2.44%
H	49	12	24.49%
I	39	3	7.69%
J	43	12	27.91%

WBC の単位は/ μ L としていたが、 $\times 1000/\mu$ L、 $10S3/\mu$ L、 $\times 10S2/\mu$ l、 \times 千/ μ L と、単位に桁数が含まれている施設、「 \times (かける)」、「x (エックス)」が含まれている施設があった。RBC や P1t でも同様の問題があった。また、尿蛋白 (定量) の単位がmg/dL ではなく、mg/L となっている施設や、尿アルブミン (1 日量) でも「mg/日」が「 μ g/mL」となっている施設もあった。

単位に桁数が表現されている場合や、単位の濃度が異なる場合があるため各施設の単位を確認して、不明の場合は、施設に問い合わせを行った上で単位変換を行っている。

4.2 検査データクレンジング及び補正

施設別の検査データの分布を視覚化することで図 2 a)が見出された。同図はHbの分布であるが、一施設だけ分布が異なり、何等かのエラーが疑われた。

調査の結果、JLAC10の転記ミスにより他の項目が誤って抽出されたことが判明し、正しいデータを抽出してデータベース登録を行った。図 2 b)は血清クレアチニンの施設ごとの分布である。各施設が同様の分布をしていることが確認できる。図 2 c)は血清カリウムの分布で、一施設の分布がやや異なっている。同施設のJLAC10の10桁目が022であることから血漿検体であり、値が10%程度低く出ている可能性が大きい。血清ナトリウムについても同様の現象がみられた。このような場合は解析目的により、必要に応じてデータに正規化の補正を施すことが可能である。図 3 に開発した検査データのクレンジングおよび補正の手順を示す。例として図 4 に図 2 c)の分布と補正後の分布を示す。図 2 d)はRBCの分布を示す。CBC関連の検査項目は標準物質の安定性の問題があり施設ごとの分布が均一ではない。またJLAC10では測定法の違いは判断できない。

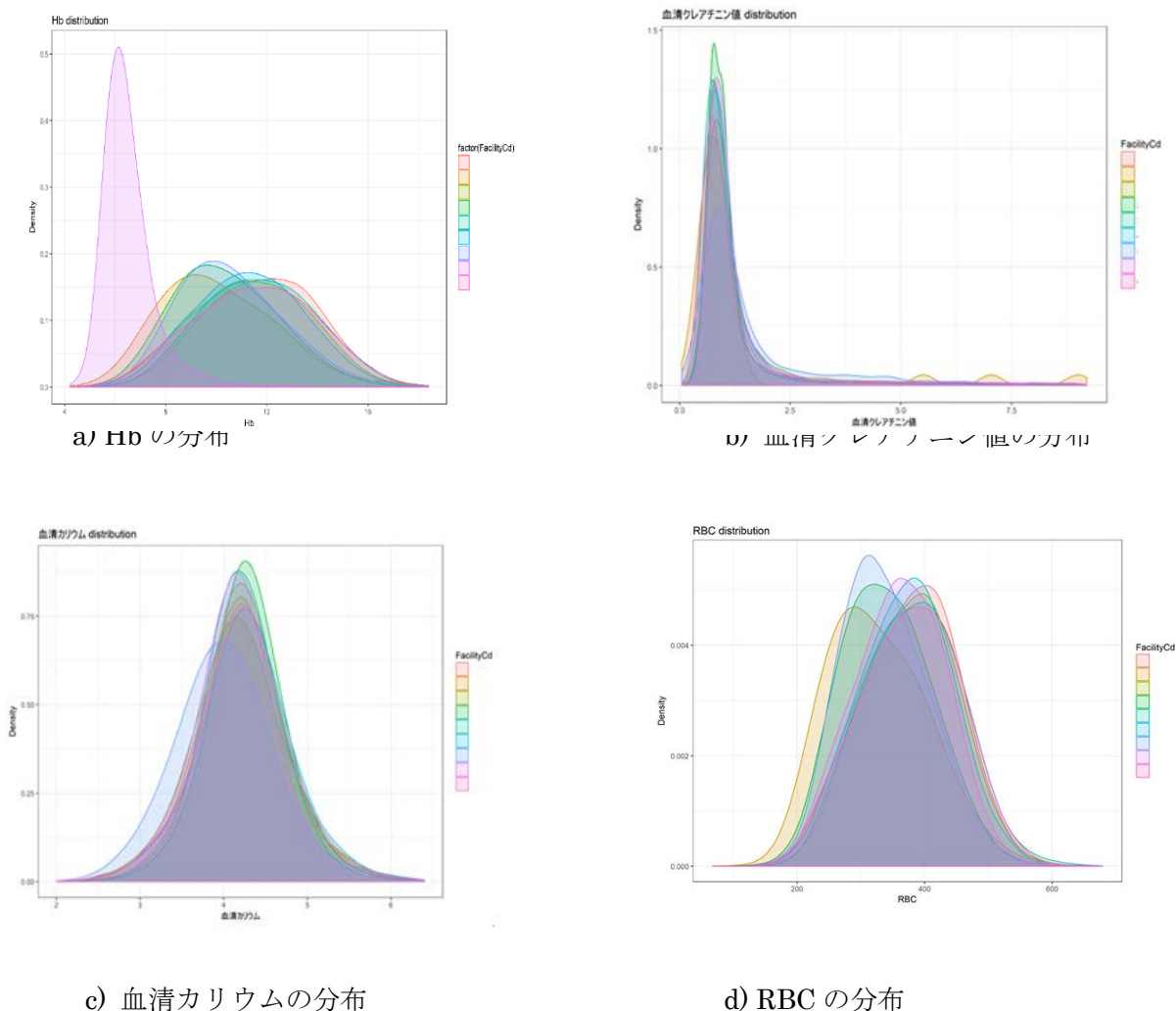


図 3 各施設の検査値分布の視覚化

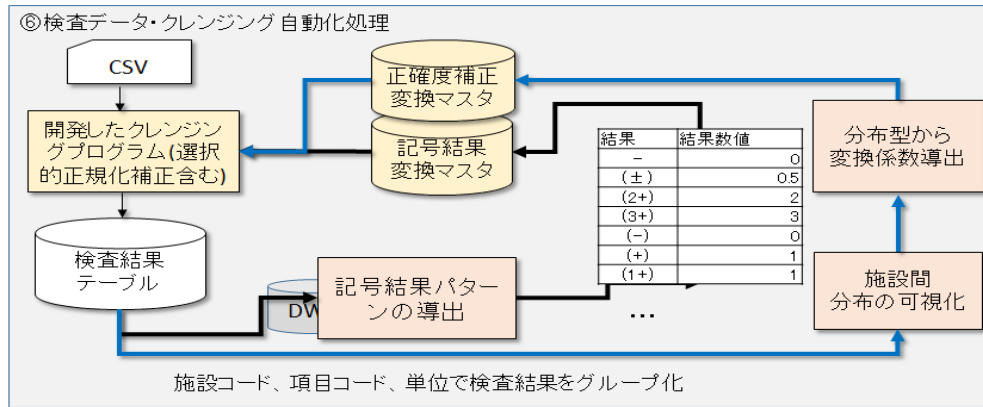


図3 検査データの可視化とクレンジング・補正

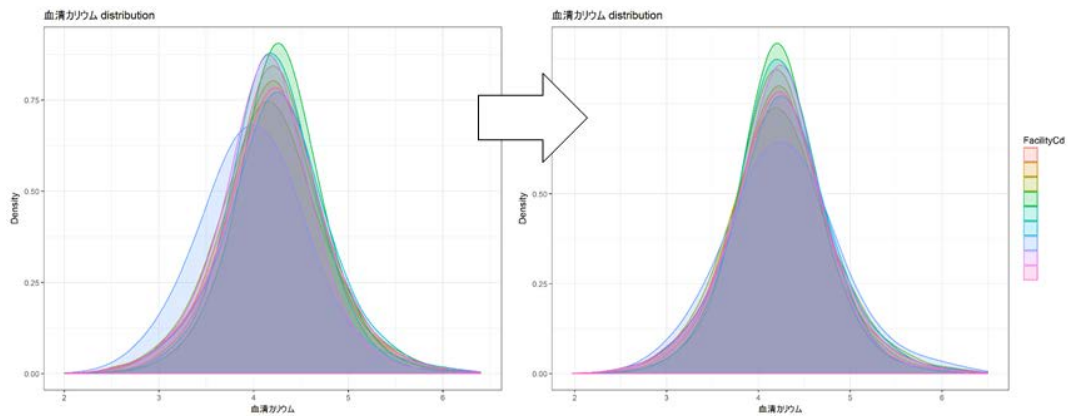


図4 図2c)の血清カリウムの分布の補正処理

5. CKDの実態の探索的調査

91861人からなるCKD-DBを対象としてCKDの実態(重症度、年齢等)の予備調査を行った。加齢とともに、男女ともにCKD重症例が増加することが判明した(図5)。

J-CKD-DBの医学研究、疫学研究における意義、発展性は極めて大きく、遵守可能で完成度の高いガイドラインの作成を可能とし、診療の質向上・均霑化に貢献できると考える。

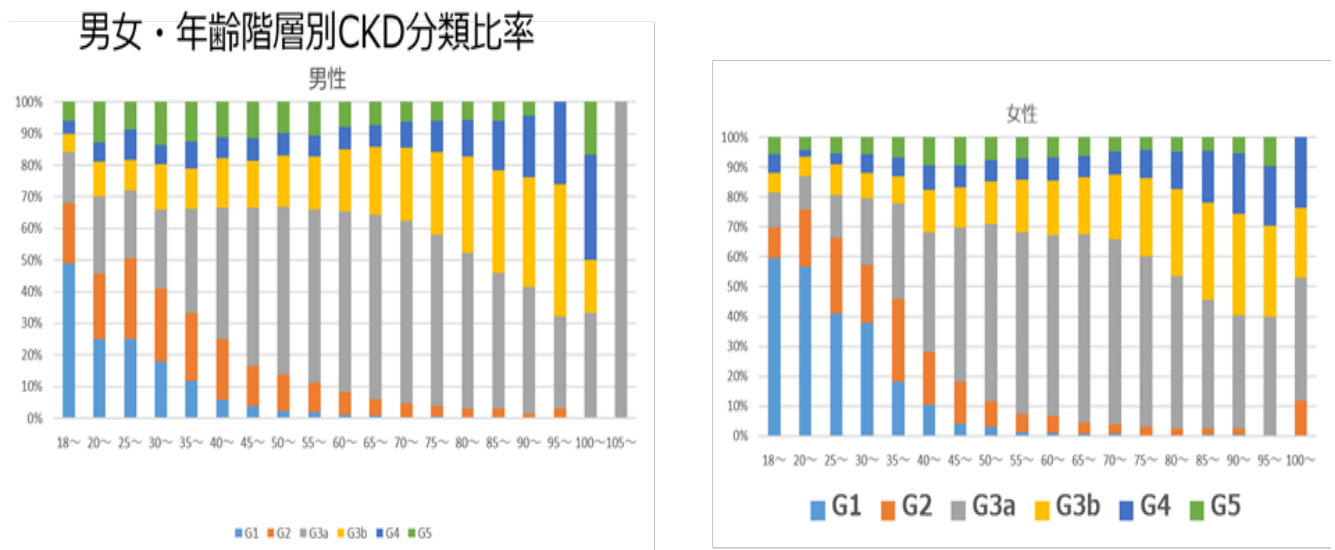


図5 男女別・年齢階層別CKD分類比率

D. 考察

SS-MIX2を用いて臨床データベースを構築する方法を開発した。

各種検査項目について施設ごとのデータ分布を可視化することで、施設間の分布の類似、相違がただちに確認できた。施設間の検査値の差異については指摘されてはいるが、どの検査項目に、どのような差異があるか、本研究により初めて明らかにすることができた。

本年度は大規模な集積データに対するクレンジング処理を重点的に行ってきた。本研究で得られた知見、臨床データベースのクオリティを高める方法は、今後の臨床データベースを用いた臨床研究・疫学研究に大きく貢献するものである。

研究活用については日本腎臓学会の倫理審査を経て承認された研究が実施されることとなるが、方法として、1) リサーチクエストを受付け、分析結果を返す方法、2) 研究目的に応じて限定された範囲のデータを提供する方法、3) 研究用データベースにアクセスしクエリーを実行できる環境を研究者に提供する方法、を検討している。2018年1月に申請された研究に対しては、1)の方法を適用して実施している。今後、2)、3)の方法について検討する。

E. 結論

本研究の目的は、ICT技術を活用して、CKDの実態調査、予後規定因子の解析、標準治療の均霈化率・準拠率の評価、有効な予防・重症化抑制策の立案、腎臓病診療の質向上、腎臓病の予後改善、透析導入率の減少、健康寿命延伸に寄与することである。本年度の主たる研究成果は以下となる。

(1) CKD該当例を自動抽出しうるアルゴリズムを作成し、電子カルテ情報(EHR: Electronic Health Record)から、SS-MIX2を活用し、多施設からCKD臨床情報を収集しDBを構築しえた。12病院から105,068人のCKD-DBを構築した。

(2) その他10病院においてもSS-MIX2標準化ストレージの整備、データ抽出作業を進めている。

(3) 得られた検査値等のデータのクレンジング・バリデーション、標準化作業を行なった。

(4) SS-MIX2を用いて臨床データベースを構築する際の課題を抽出し、解決方法の開発に着手した。この成果は他研究にも活用できるものとなる。

(5) 91861人を対象として、本邦のCKDの実態についての探索的解析を行った。

本研究の成果より、J-CKD-DBの医学研究、疫学研究における意義、発展性は極めて大きく、遵守可能で完成度の高いガイドラインの作成を可能とし、診療の質向上・均霈化に貢献できると考える。また、本研究で得られた臨床データベースのクオリティに関する知見、およびこれを高める方法は、今後の臨床データベースを用いた臨床研究・疫学研究に大きく貢献するものである。

F. 健康危険情報

特に無し

G. 研究発表

1. 論文発表

- 1) Wang J, Zhang L, Tang SC, Kashihara N, Kim YS, Togtokh A, Yang CW, Zhao MH; ISN North and East Asia Regional Board. Disease burden and challenges of chronic kidney disease in North and East Asia. *Kidney Int*;2018
- 2) 柏原直樹、桑原篤憲、長洲一、岡田美保子：包括的慢性腎臓病データベース (J-CKD-DB) . 日本腎臓学会誌2017; 59(7):1034-1041
- 3) 桑原 篤憲、岡田 美保子、片岡 浩巳、柏原直樹、J-CKD-DB 企画運営委員会：電子カルテ情報を活用した多施設臨床効果データベースの構築と課題、*医療情報学* 2017; 37 (Suppl.): 996-999
- 4) 山本陵平、新澤真紀、守山敏樹、猪阪善隆、杉山斉、丸山彰一、中島直樹、岡田浩一、岡田美保子、柏原直樹：コホート研究を目的とした情報処理システム：コホートメーカー、*医療情報学*2016; 36 (Suppl.): 544-545
- 5) 桑原篤憲、片岡浩巳、柏原直樹、岡田美保子：病院情報システムとMCDRSを用いた多施設共同臨床データベースの構築とその課題-日本における慢性腎臓病(Chronic Kidney Disease: CKD)の包括的臨床データベース(J-CKD-DB) . *医療情報学*2016; 36 (Suppl.): 1126-1127

2. 学会発表

- 1) 柏原直樹:疾病克服を目指した日本腎臓学会の取り組み. 第 40 回日本高血圧学会総会,2017/10/20,国内
- 2) 柏原直樹:腎臓病克服への挑戦~日本腎臓学会の取り組み~. 第 62 回日本透析医学会学術集会,2017/6/16,国内

- 3) 柏原直樹:臨床研究等 ICT 基盤構築研究事業腎臓病データベースの拡充・連携強化と包括的データベースの構築.第60回日本腎臓学会学術総会,2017/5/26,国内
- 4)

H. 知的財産権の出願・登録状況

(予定を含む。)

特に無し