

課題番号 : 23指102
研究課題名 : 病院情報システム (MegaOak) からの医療情報の自動取り込みと電子化診療情報のデータベース化に関する研究
主任研究者名 : 菊池 嘉
分担研究者名 : 星野 隆之、川崎 敏克
キーワード : 情報工学、医療・福祉、病院情報システム、電子カルテ、データベース
研究成果 : ACC 患者データベースの構築、改修・発展

3年の研究にあたり、当初は、当センターの病院に導入されている電子カルテシステム MegaOak から、各種検体検査、放射線検査、生理検査の依頼情報などを引き出し、ある疾患の特異的なデータベースを構築することを目指すこととして、主任研究者の所属科である、エイズ治療研究開発センター (ACC) を手始めとして、患者データベースの構築を基本目標とした。

この基本目標が達成できたら、下記の3つの要点を順次達成することが本課題の発展到達目標であった。その3要点とは、

- (1) 臨床各科、特に少人数の担当医師からなる診療科からの診療情報の取り出しへ協力する
 - (2) 臨床評価指標である Quality Indicator (Clinical Indicator) へ対応する
 - (3) 将来の CDISC (Clinical Data Interchange Standards Consortium) 化への対応の足掛かりとする
- ことであった。

本研究に着手する前年の平成22年8月に、病院電子カルテシステムが、現行の MegaOak に移行した。それまでは、Leaf と呼ばれるオーダーリングシステム (電子カルテではなく単に検査データのオーダーを入力し結果を参照するシステム) による運用で、採血検査データ、処方データなどが診療の補助ツールとして使用されていた。このオーダーリングシステムである Leaf から電子カルテシステム MegaOak へ採血検査データ、処方データなどの多くのデータが移行され診療現場での過去のデータを利用することは可能であった。しかしながら、電子カルテには、研究者が頻用するエクセル形式などとは全く異なったデータ保管形式で格納されている。電子カルテ画面上では、日々のデータが一覧可能であるが、それを取り出して臨床研究に役立てることは容易ではなかった。

さらに、一人の患者データではなく、同一疾患の患者データ群を比較検討することは更に困難を極めた。

そこで、病院電子カルテシステムに格納されている医療情報を、エクセルなどのデータ解析に用いるために人手による入力作業をかけずに、電子媒体に取り出して何とか利活用できないかという、発想から本研究は開始された。

初年度に、電子カルテシステムからの匿名化されないデータの引き出しは技術的に可能であることがわかったが、匿名化されないままのデータの引き出しは、情報漏洩にも繋がり、個人情報への取扱いの点で問題があるので、次年度へかけて、データの匿名化の作業に着手した。

ACC では 1997 年 4 月に当センター内に設立されて以来の独自のデータベースがあり、日々の新患患者の ID を手打ち入力で脈々と保持してきた患者リストが存在していた。このデータを利用して、電子カルテ側に ID を要求し、その ID に紐付いたデータを引き出すことができると、そのデータを研究者が解析しやすい形に変形して応用することができる。引き出す ID を暗号匿名化することにより、個人情報秘匿されて、研究に活用可能となった。

最初は ID と幾つかのデータのみが紐付けされて取り出し可能となったが、徐々に項目を増やして、多くのデータが取り出せる仕組みが整っていった。検査データは、各患者の ID と日付と検査項目、検査結果のセットで一つの意味を持った集合となる。この集合体を集めて、暗号化して引き出しながら、目標とする研究用 ID を付与したデータベースの構築が完成した。

データベースを構築した後は、各患者の日々のデータと 2 年度目以降から薬剤リストも引き出し可能となり、薬剤の処方日と検査データが一体となったデータベースの完成形ができあがった。

ACC のデータベースは完成し、検査データがある異常値を超える場合の ID を検索可能な機能も整えることができた。残念ながら、期間を限定してある薬剤を使用して、その際にあるデータが異常を来した症例までを同定する仕組みの確立までは本研究には為しえなかったが、簡単な検索は可能となった。

臨床医は、患者さんの実名で病気の状態を把握しているので、検索可能な状況となっても暗号化された ID では誰のものなのか全くピンと来ない。病院電子カルテと同様に閉ざされたネットワークを構築し、その範囲内で、セキュリティーを保ちながら、暗号化匿名化された情報を復号化し、自科で整えていた患者リストに突合することにより、検索可能なデータベースを構築することも進めた。

この構築により、2 カ年目以降、毎週データをアップデートするところまで機能が充実した。このアップデート作業は一作業ごとに費用が発生するとことになるが、費用面さえ気にしなければ、連日でもデータを更新することも技術的には可能となった。

処方履歴の取り込みが可能となった後は、初回治療導入時の処方内容を表示する新たなツール開発まで達成した。登録された各患者の初回導入時の、抗ウイルス剤の治療組み合わせが一覧できることは、長期にわたる HIV 治療の基本となるものであり、とても意義深いことであった。この仕組みを応用することで、直近の処方履歴から、現在使用されている薬剤の組み合わせなども処方箋ベースで表示できることが可能となり、閲覧ソフトを整えれば、表示が可能となった（費用の関係で実装には至らなかった）。

本研究で整えた、実臨床に用いている電子カルテのデータから、該当する ID を電子カルテに投げかけて、その ID に付随する検査結果、処方内容などを引き出し、電子カルテから取り出す際には暗号化され、特に患者 ID などを匿名化して取り出し、データベースへ移行させ、必要に応じて復号化する仕組みを、各科の要望に応じて対応することで、臨床研究への応用はもちろん、日々の業務実績の把握などにも役立てる道が開けた。

臨床各科に応用するには、各科が必要なデータセットが異なるため、それなりの経費はかかるが、それをすこしずつ広げることで、センター全体の発展に貢献できる可能性がある。各種のデータを抽出することができるため、そのデータに基づいた臨床評価指標（QI もしくは CI）を用いている場合は、表出もかろうとなる道が開けた。簡単なものであれば、月次で更新されることへの応用も期待でき、病院全体への貢献度はさらに高まった。

研究開始当初は、CDISC (Clinical Data Interchange Standards Consortium) 開発には、1 治験で数千万円（簡単なものでも 2 千万円程度）の費用がかかるといわれていたが、この研究によるデータの引き出しが可能となったことで、CDISC に準拠したデータ取り出しにも何とか間にあう状況まで整えることができた。幸いこの研究期間中に CDISC でデータを抽出することを要求される海外治験がなかったため応用には至らなかったが、別予算も必要とされることなく研究期間を終えた。

CDISC に対応する代わりに、治験管理の際に必要な、SDV (Source Data Verification) のための電子カルテの参照画面を確立した。特定の患者 ID のみを治験管理室で設定して、治験依頼者はその一覧リストからしか患者を選択出来ず、無用な患者の閲覧制限ができる電子カルテの閲覧設定を作成した。また電子カルテの全般が見渡せるわけではなく、治験に必要な部分のみ、即ち、電子カルテの記事記載欄、検査結果一覧、退院サマリー画面、その他の必要な画面のみの参照ができるように整えることができた。

3年前から限られた予算で、当時は導入が困難であった SS-MIX2 などの高価なデータストレージを使わずに、院内における臨床データを利活用することができたことで、本研究の意義は果たせたと思われる。しかしながら、まだ全科に渡り汎用性があるわけではなく、今後も発展の余地は残った。CDISC の要求がなかったことも幸いしたが、まだ簡易なデータの利活用には道は遠く、今後も努力が必要な領域であると感じた。

ACC のデータベースがほぼ完成したため、ACC に関連した多くのコホート研究が可能となり、多くの英語論文が排出されるようになった。主な論文と、学会発表の一部を別掲した。

Subject No. : H23-designated-102 23 指 102

Title : Studies on automatic ETL (extract, transform and load) of clinical information from hospital information systems (MegaOak) and creating a research database from Electronic Medical Records (EMR)

Researchers : Y. Kikuchi T. Hoshino, T. Kawaski

Key word : hospital information systems, Electronic Medical Records (EMR), database

Abstract :

Aim of the Project: This study aimed to extract various clinical information from MegaOak, NCGM hospital's Electronic Medical Records (EMR) such as specimen exam results, radiology results and exam order information according to various diseases and to create disease specific database, starting from the project leader's department (ACC). With this original goal, the study group intended spreading to the hospital-wide program, following three items subsequently:

- I. Data extraction cooperation for various clinical departments in the hospital, especially departments with less manpower.
- II. Response to the clinical evaluation indicator, Quality Indicator (Clinical Indicator).
- III. Establishing the basis for future implementation of Clinical Data Interchange Standards Consortium (CDISC).

Project Summary: The hospital-wide EMR, MegaOak was installed in August 2010, several months before this project started, taking over the old ordering system, Leaf, along with other clinical assisting tools for blood test results and prescriptions, whose past data could be reviewable by clinicians. Contrary in the EMR, clinical data are stored in a completely different format from desktop friendly interface such as Microsoft Excel, which made it impossible to extract the data of interest for clinical research although the everyday data are displayed comprehensively on EMR browser. Furthermore, it was extremely difficult to compare data of a group of patients with the same disease, even harder than data from individual patients.

So, the project implementation started with the intention to extract the clinical data stored in EMR and transfer into more common electric media to utilize as such for Excel analysis without manual, human data entry. The first year results proved that it was technically doable to pull out identifiable data from EMR, which would be problematic from the viewpoint of protected health information. The second year was implemented for de-identification of the data.

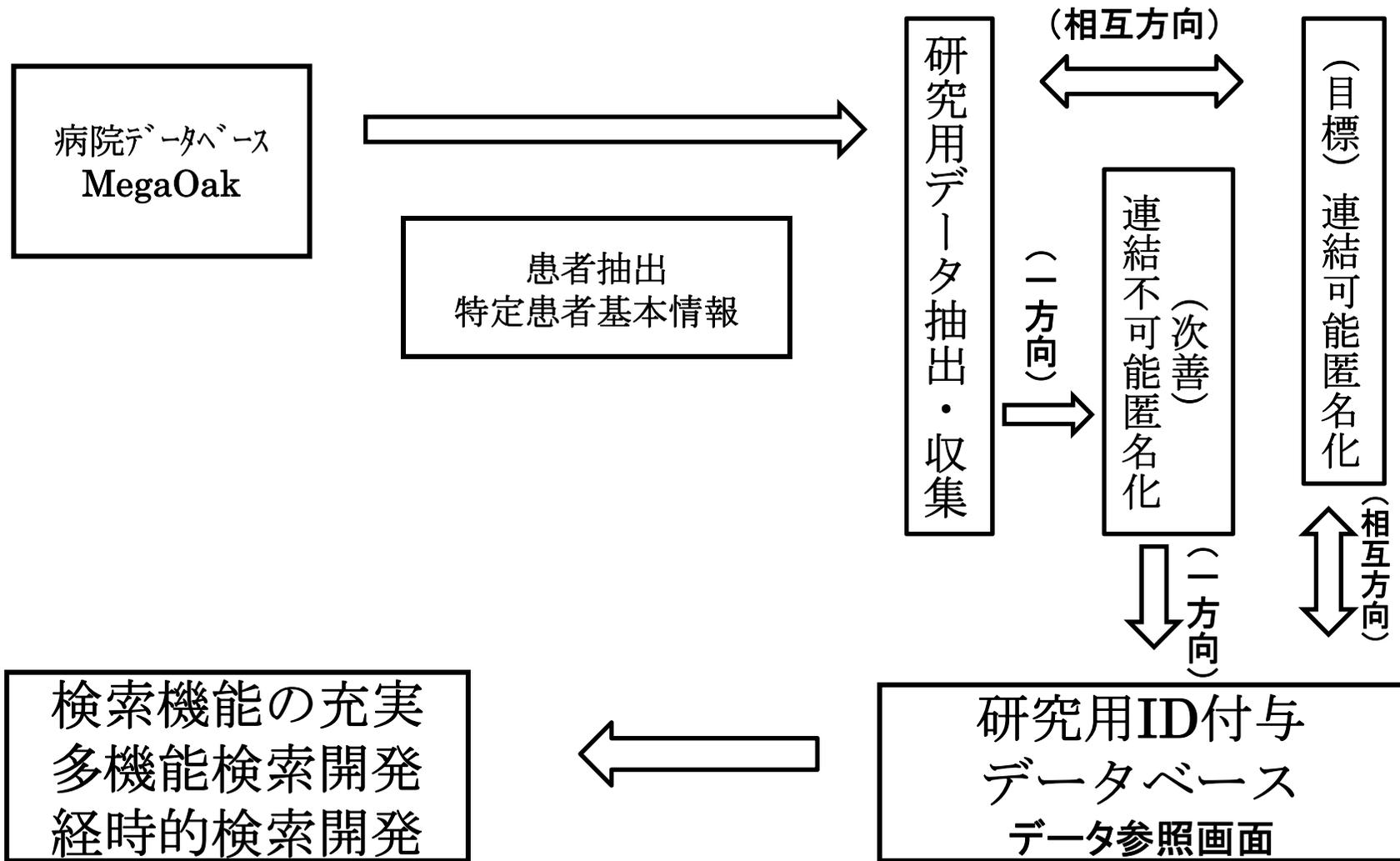
In ACC, the IDs for the clinical database have been assigned and entered manually since ACC was established in April 1997, yielding a list of IDs for the patients which was integral for de-identified data search, extraction and utilization for research purposes. The second year the study group enabled data updating functions on a weekly base. Though this function would cost more for daily updates, we successfully developed the operational capacity, according to which daily updates costs merely five times more per week. But it is a significant improvement in clinical values to be able to access to the latest clinical data every day by the daily update function. By the third and last year of the project, establishment of ACC database was almost completed, weekly updates are performed, not only transfer of clinical data as it used to be, but also prescriptions for medications are wholly transferred, and we were even able to achieve to develop a new tool to indicate especially the clients' historical regimen at the initiation of the first treatment within the year. Ongoing for the next year, the study group would like to apply and create databases for other diseases, especially specific diseases of health policy importance to open possibilities for clinical study application and to analyze everyday operational outcomes, thus contribute to NCGM as whole to achieve its goals. Also, because this study makes it possible to extract more various kinds of

Researchers には、分担研究者を記載する。

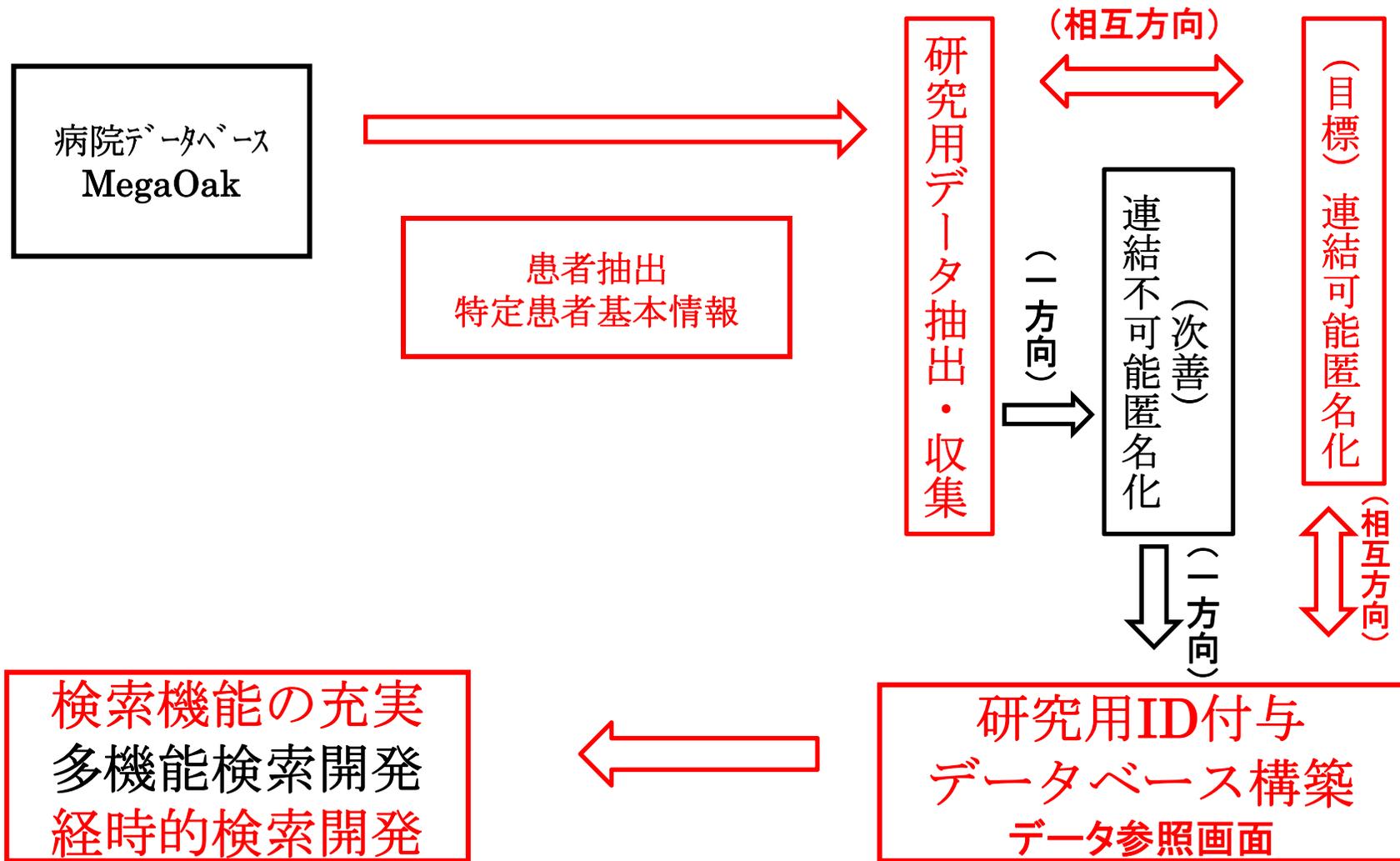
hospital data, these databases can produce monthly updates of clinical evaluation indicators (QI or CI) in order to meet operational goals of the NCGM hospital even further. In clinical trials in the Western countries, hospital data are extracted as whole by CDISC for quality assured data collection methods in order to meet FDA's international standards. However in Japan, there are yet few facilities that achieved adaptation to CDISC and NCGM hospital too has not been operating the already installed CDISC systems. Because some of Japanese clinical information, tests and prescriptions meet EMS industrial standards and others do not, it is extremely cumbersome and difficult to extract quality data automatically without depending on human hands. The initial cost estimation for CDISC installation was in the neighborhood of 20 millions yen but we were able to save those budgets by this project. Collaborative efforts by multiple facilities are expected to be clinically more valuable, but in such cases, such efforts might become highly problematic with different systems disagreeing with each other. So, in the last year, the study group developed system for SDV (Source Data Verification, utilized in various clinical trials) instead of CDISC.

As the result, creation of ACC database enabled numerous ACC related cohort studies and research papers in English; we attached a partial list of those publications and academic conference presentations.

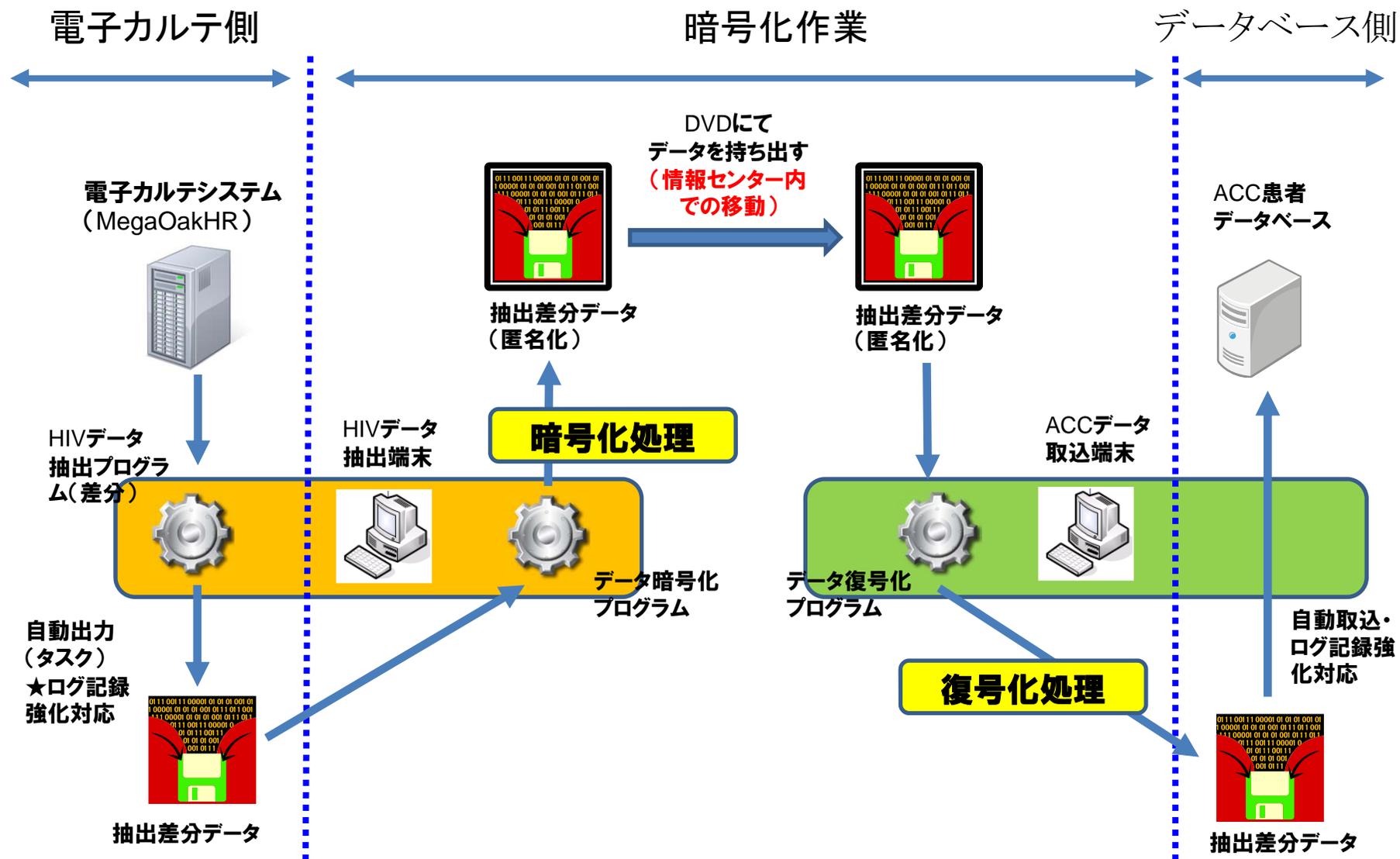
23指102菊池班
病院情報システム(MegaOak)からの医療情報の自動取り込みと電子化診療情報のデータベース化に関する研究 (申請当初の概略図)



23指102菊池班
病院情報システム(MegaOak)からの医療情報の自動取り込みと電子化診療情報のデータベース化に関する研究 (完成した部分)



23指102菊池班
 病院情報システム(MegaOak)からの医療情報の自動取り込みと電子化診療情報のデータベース化に関する研究(研究終了時データ移行分までの概略図)



23指102菊池班

病院情報システム(MegaOak)からの医療情報の自動取り込みと電子化診療情報のデータベース化に関する研究(データベース参照画面)

患者情報入力 - Microsoft Access

患者ID: 1000001 | 患者名: タミー 2 | 更新日: 2013/07/23 15:53:37

よみがな: タミーニ | 生年月日: 1999/01/01 | 年齢: 14 | 性別: 男

住所: | A-net: 保留 | 検体保存: 未

初診日: 2000/01/01 | 外来医: | 医師研ID: | 担当CN: | 身長: 170

備考: 本番確認用に使用させていただいております。

check
 カルテ(外)
 更新ナン

HA-Ab 陽性
 HBeAg 陽性
 HBeAb 陽性
 HCV-Ab 陽性
 CMV-Ab 陽性
 HBeAg 陽性
 HBeAb 陽性
 Taxo 陽性
 Ameba 陽性
 HBsAb 陽性
 HBcAb 陽性
 TPHA 陽性
 備考 あり

日々データ | 感染症データ | 入院データ

日時	外来ID	外来内容	病名ID	日本語病名	疾患詳細	除	WBC	Neut	Neu/E	Lymph	EO	CD4%	CD4手	CD8%	CD8手	4/8	CD4	CD8	VL(定)	VL(推)	RBC	Hb	MCV	Plt	Ab	Tybl	GOT	GPT	LDH	ALP	γ-GT		
2012/11/07	9	その他の特	200	耐性遺伝子																													
2012/09/30	2	外来データ					1	2	0																								
2012/01/30	2	外来データ					1	2	0	3	4	5		6	0.83	0	0	7	8	9	10												
2010/04/10	2	外来データ	12	口腔カンジダ	(L,N,M,I)																												
2009/02/03	2	外来データ																															
2006/07/18	2	外来データ	16	カブツコウシ腫	1		3300	3	99	500	4	40		4		10	6600	660	2	2	22	22	22	22	22	22	4	4	4	4	4	4	
2002/02/02	2	外来データ	12	口腔カンジダ	あああ		1250	22	275	300	3	20	333	10		2	750	375	4.5	3	22	22	3	22									
2002/01/01	2	外来データ	10	カリニ肺炎	あああ		2200	3	66	500	4	50		4		12.5	5500	440	4.1	4	22	4	4	4	4	4	4	4	4	4	4	4	
2001/05/17	2	外来データ					1400	43	602	300	4	25		6		4.17	1050	252			4	4	4	4	4	4	4	4	4	4	4	4	

レコード: 1 / 9 | 検索

ATOK 高速 R漢 読 兼小

11:01 2013/07/24

課題番号 : 23指102

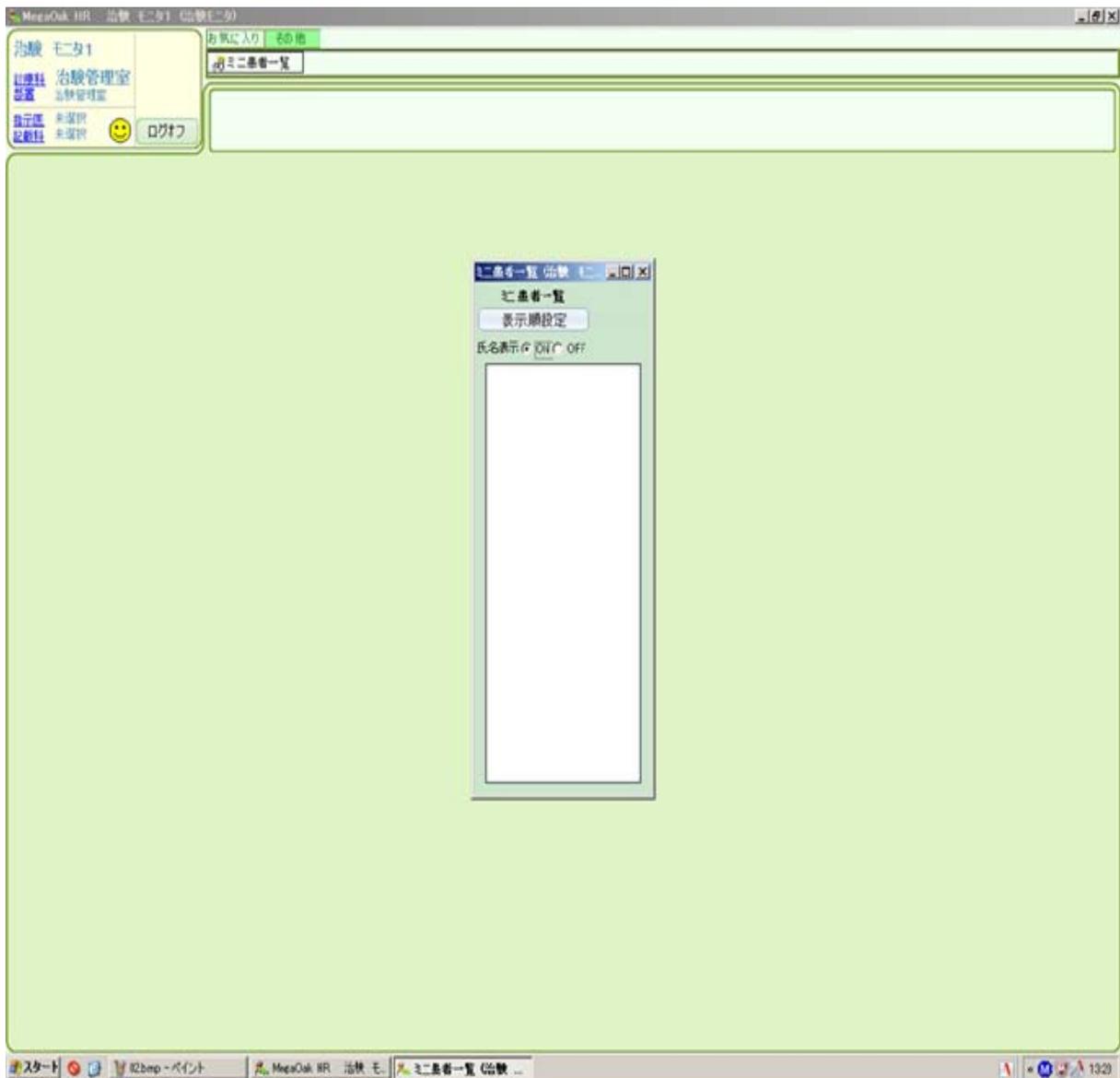
研究課題名 : 様々な治験に関する診療データの病院情報システムからの抽出の検討

主任研究者名 : 菊池 嘉

分担研究者名 : 川崎 敏克

キーワード : 病院情報システム、電子カルテ、データベース

研究成果 : 主任が行った、医療情報の自動取り込みと電子化診療情報のデータベース化に関する検討の一貫として、治験の際に必要な原資料との照合検証の際に必要な SDV (Source Data Verification) のための電子カルテの参照画面を確立した。



研究開始時には CDISC (Clinical Data Interchange Standards Consortium) に対応出来るデータ抽出の準備を整える算段で本研究に参画したが、過去3年間に本邦においては CDISC の要請が高まらず、CDISC の方の検討は進展しなかった。最終年度の今年度も切望していたが依頼はなかった。しかしながら、3か年の研究期間においては、国内で行われた CDISC の紹介セミナーに参加したほか、Web上で CDISC の概要を定期的に調査するなどして、国内外における流れに追従すべく努力を払った。このため、最終年度は、SDV のための電子カルテの参照画面の設定にテーマをシフトして、治験管理室長としての役割を果たした所となった。

SDV 画面の設定

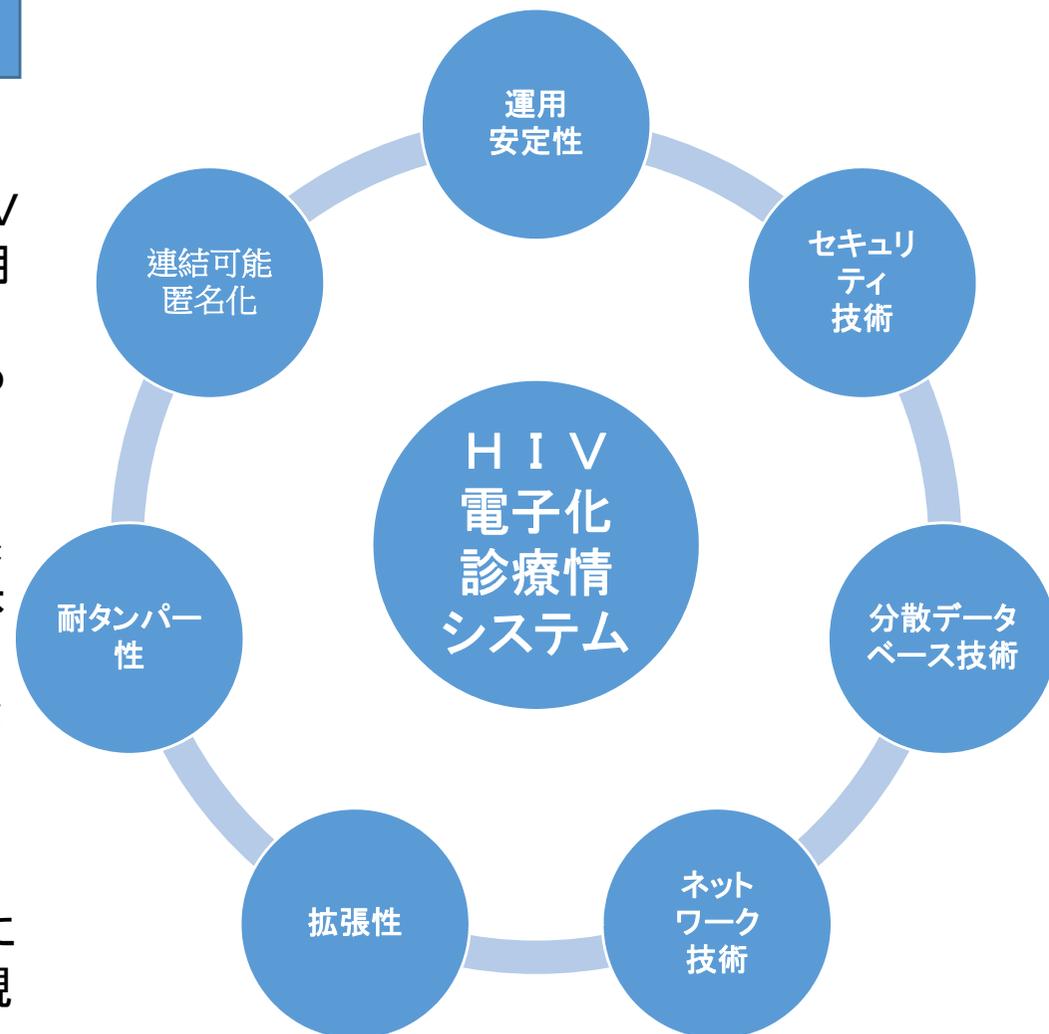
治験依頼者が電子カルテに通常にアクセスすると、入院一覧、外来一覧画面などから、治験に関係のない患者全てにアクセスできてしまう状態であった。このため、特定の患者 ID のみを治験管理室で設定して、治験依頼者はその一覧リストからしか患者を選択出来ず、無用な患者の閲覧制限ができる電子カルテの閲覧設定を作成した。また電子カルテの全般が見渡せるわけではなく、治験に必要な部分のみ、即ち、電子カルテの記事記載欄、検査結果一覧、退院サマリー画面、その他の必要な画面のみの参照ができるように整えることができた。具体的には、治験管理室の職員 ID で電子カルテにアクセスして、治験画面を開き、そこで患者登録画面から参加者 ID 登録することによって、SDV 画面では治験参加者のみのカルテ閲覧が可能となった。

23指102菊池班（分担研究者 星野 貴之）
電子化診療情報システムの安定運用に必要なセキュリティ
技術及び分散データベース技術の情報工学的検討 その1

安定運用のための技術的基盤整備

電子カルテシステムMega-Oakより、HIV陽性の患者を抽出し、連結可能匿名化した新たな研究用IDを付与し、HIV診療に必要なデータを自動抽出するシステムの安定運用において必要とされるセキュリティ技術、分散データベース技術、ネットワーク技術等に関して技術的な検討を行った。

特に、電磁的に取り出したデータをもとに、各患者の時系列表示ができるデータベースを構築し、さらに異常値を示したデータなどを簡単な操作で検索可能な機能も備えたシステムの安定運用に必要なIT技術に関して、障害の発生のおよび、障害発生時の回復時間の短さといった可用性と、非正規な手段による機密データの読み取りを防ぐ能力である耐タンパー性と、今後の発展性を保証する拡張性とを中心に検討を行い当該システムの運用安定性に関して最適な技術的手段の選択を行い、安定運用を実現するための技術的基盤整備を行った。



23指102菊池班（分担研究者 星野 貴之）
電子化診療情報システムの安定運用に必要なセキュリティ
技術及び分散データベース技術の情報工学的検討 その2

Layer毎の設計

CSV: 保守が容易・習熟した技術者が多い

国際医療研究センター
電子カルテシステム
Mega-Oak(NEC社製)



データ抽出
Layer

匿名化機能

- アルゴリズム
アタッチメント
- 連結可能
- 連結不可能

形式変換

- 元システムのバージョンアップに対し、このLayerの設定変更で対応可能

データ蓄積
管理Layer

暗号化機能

- O DSA
- O AES
- X DES

変換
Layer

各システム毎の出力
定義

H I V
電子化
診療情
システム

他の研究機関



HL7:国際規格(将来構想)

他
システム

参考文献

「医療情報システムの安全管理に関するガイドライン第4.1版」厚生労働省、平成22年2月

研究発表及び特許取得報告について

課題番号： 23指102

研究課題名： 病院情報システム (MegaOak) からの医療情報の自動取り込みと電子化診療情報のデータベース化に関

主任研究者名： 菊池 嘉

論文発表

論文タイトル	著者	掲載誌	掲載号	年
Ritonavir-boosted darunavir is rarely associated with nephrolithiasis compared with ritonavir-boosted atazanavir in HIV-infected patients.	Nishijima T, Hamada Y, Watanabe K, Komatsu H, Kinai E, Tsukada K, Teruya K, Gatanaga H, Kikuchi Y, Oka S.	PLoS One	8(10):e77268	2013年
Is ritonavir-boosted atazanavir a risk for cholelithiasis compared to other protease inhibitors?	Hamada Y, Nishijima T, Komatsu H, Teruya K, Gatanaga H, Kikuchi Y, Oka S.	PLoS One	8(7):e69845	2013年
Preemptive therapy prevents cytomegalovirus end-organ disease in treatment-naïve patients with advanced HIV-1 infection in the HAART era.	Mizushima D, Nishijima T, Gatanaga H, Tsukada K, Teruya K, Kikuchi Y, Oka S.	PLoS One	8(5):e65348	2013年
Raltegravir can be used safely in HIV-1-infected patients treated with warfarin.	Honda H, Gatanaga H, Aoki T, Watanabe K, Yazaki H, Tanuma J, Tsukada K, Honda M, Teruya K, Kikuchi Y, Oka S.	Int J STD AIDS.	23(12):903-4	2012年

学会発表

タイトル	発表者	学会名	場所	年月
当センターにおける初回抗HIV療法の動向	塚田訓久、水島大輔、西島健、青木孝弘、木内英、渡辺恒二、本田元人、矢崎博久、田沼順子、瀧永博之、照屋勝治、菊池嘉、岡 慎一	日本エイズ学会学術集会・総会	熊本	2013年

研究発表及び特許取得報告について

当センターにおいて初回抗HIV療法の際に 選択された抗HIV薬の変遷	塚田訓久、橋本 亜希、矢崎博 久、水島大輔、 西島 健、小林 泰一郎、青木孝 弘、渡辺恒二、 木内 英、本田 元人、田沼順 子、照屋勝治、 潟永博之、菊池 嘉、岡 慎一	日本エイズ学会学 術集会・総会	横浜	2012年
---------------------------------------	--	--------------------	----	-------

その他発表(雑誌、テレビ、ラジオ等)

タイトル	発表者	発表先	場所	年月日
なし				

特許取得状況について ※出願申請中のものは()記載のこと。

発明名称	登録番号	特許権者(申請者) (共願は全記載)	登録日(申請日)	出願国
なし				

※該当がない項目の欄には「該当なし」と記載のこと。

※主任研究者が班全員分の内容を記載のこと。