

厚生労働科学研究費補助金

政策科学総合研究事業

(臨床研究等ICT基盤構築・人工知能実装研究事業)

電子カルテ情報をセマンティクス(意味・内容)の標準化により分析可能なデータに変換するための研究

平成29年度 総括研究報告書

研究代表者 宮本 恵宏

平成30(2018)年 3月

目 次

I . 総括研究報告

電子カルテ情報をセマンティクス（意味・内容）の標準化により分析可能なデータに変換するための研究	-----	1
宮本 恵宏		

II . 分担研究報告

1	自然言語処理によるMACEイベント捕捉の検討に関する研究	-----	6
	西村 邦宏・中村 文明・竹上 未紗・安田 聡・宮本 恵宏		
2	臨床効果データベース事業におけるデータ出力方法の検討に関する研究	-----	8
	安田 聡・斉藤 能彦		
3	データ転送プログラムによるデータ収集に関する研究	-----	9
	中山 雅晴		
4	SS-MIX2データ抽出システム実装とテストおよびアウトカムテンプレート仕様検討	--	11
	的場 哲哉		
5	臨床効果データベース事業の学会支援	-----	12
	小室 一成		
6	自然言語解析による診断名判断システムの開発	-----	13
	宍戸 稔聡・平松 治彦・上村 幸司・竹村 匡正		
7	SS-MIX2を利用した臨床効果データベースのデータ収集に関する研究	-----	16
	宇宿 功市郎・辻田 賢一		

III . 研究成果の刊行に関する一覧表	-----	17
----------------------	-------	----

平成 29 年度厚生労働科学研究費補助金

(政策科学総合研究事業(臨床研究等ICT基盤構・人工知能実装構築研究事業)) 総括研究報告書

「電子カルテ情報をセマンティクス(意味・内容)の標準化により分析可能な
データに変換するための研究」

研究代表者	宮本 恵宏	国立循環器病研究センター・循環器病統合センター・センター長
分担研究者	竹村 匡正	兵庫県立大学大学院・応用情報科学研究科・教授
	中村 文明	国立循環器病研究センター・循環器病統合センター・室長
	竹上 未紗	国立循環器病研究センター・研究開発基盤センター 予防医学・疫学情報部 室長
	興梠 貴英	自治医科大学・医療情報部・教授
	中山 雅晴	東北大学大学院・医学系研究科・教授
	的場 哲哉	九州大学病院・循環器内科・講師
	小室 一成	東京大学大学院医学系研究科・循環器内科・教授
	斎藤 能彦	奈良県立医科大学・循環器内科・教授
	安田 聡	国立循環器病研究センター・副院長・心臓血管内科部門長
	穴戸 稔聡	国立循環器病研究センター・研究推進支援部・部長
	西村 邦宏	国立循環器病研究センター・循環器病統合情報センター・室長
	平松 治彦	国立循環器病研究センター・情報統括部・部長
	上村 幸司	国立循環器病研究センター・研究推進支援部・室長
	辻田 賢一	熊本大学大学院・生命科学研究部・教授
	宇宿 功市郎	熊本大学医学部附属病院・医療情報経営企画部・教授
研究協力者	住田 陽子	国立循環器病研究センター・循環器病統合センター・専門職
	都島 健介	東京大学医学部附属病院・循環器内科・助教
	中村 太志	熊本大学医学部附属病院・医療情報経営企画部・副部長

研究要旨

本研究では、電子カルテの記事情報から自然言語処理を活用して自動的に MACE であると判断するためのシステムを開発し、電子カルテ情報を用いた MACE のビッグデータ分析を行うためのシステムを開発する。日本語で記述される電子カルテからの臨床データベースにおいては初めての試みである。昨年度までに、電子カルテの標準フォーマットである SS-MIX2 の整備を行った。日本循環器学会標準フォーマット(SEAMAT)に基づき、心電図、心エコー、心臓カテーテル検査の結果を SS-MIX2 拡張ストレージに格納する作業を行った。また、国立循環器病研究センターにおいて電子カルテ記事の抽出を行い、電子カルテ記事の自然言語処理を行う準備を行った。本年度(平成 29 年度)は、電子カルテシステム上のデータの確認と、機械学習を用いた症状記載の自動抽出に関する実験、自然言語処理を行う準備である医療用語辞書の作成、SS-MIX2 データからのデータベースを構築と電子カルテ記事の抽出を行った。

A. 研究目的

高齢化社会の中にある我が国をはじめとする先進諸国において、循環器疾患が急増している。循環器疾患は再発を繰り返し徐々に進行していくという臨床経過をたどることが多い。例えば、虚血性

心疾患では再発・入院を繰り返して終末像として心不全を呈することがしばしばある。そのため循環器疾患においては、Major Adverse Cardiac Event (MACE) とよばれる主要有害心血管イベントを発生させないための再発予防が大事である。循環器疾患

の新規治療法の開発目標として、MACEの発生減少を目標としたものを開発することも考えられるが、MACEを判断するためには担当した臨床医の判断が診療録を読み返し判断するしかない。そのため、レセプト/DPCなどの診療報酬請求情報を使用した分析、または電子カルテ情報を用いてビッグデータの分析においては、MACEなどのイベントをアウトカムにした研究をすることができないという限界がある。本研究では、電子カルテの記事情報から自然言語処理を活用して自動的にMACEであると判断するためのシステムを開発し、電子カルテ情報を用いたMACEのビッグデータ分析を行うためのシステムを開発する。日本語で記述される電子カルテからの臨床データベースにおいては初めての試みである。日本循環器学会の事業で実施している医療コストがかかる疾患・治療（心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法（CRT））と循環器領域で特にその重要性が指摘されている疾患（急性心不全など）を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである「臨床効果データベース」を用いて自然言語解析による診断判定システムの構築をおこなうことを目的としている。

B. 研究方法

疾患コホート研究であり、虚血性心疾患、急性心不全の患者を対象とする。対象施設は、国立循環器病研究センター、東京大学、自治医科大学、自治医科大学さいたま医療センター、東北大学、九州大学にてデータの収集を行う。臨床効果データベースから、患者基本情報、診断名、入退院情報、経時的な内服薬、経時的な臨床検査情報、経時的な生理検査情報、経時的な心臓カテーテル検査情報を取得する。別途、電子カルテの記事情報を、臨床効果データベースと同じ匿名化番号にて匿名化したIDにて連結可能匿名化して受け取り、臨床効果データベースのデータと連結を行うことにより、電子カルテの記事情報と臨床データの結合を行う。さらに、入退院情報、検査結果等のデータからMACEによる入院かどうかの判断を行う。これらのMACEのそれぞれのイベントに対して、ICD10を用いて病名のコーディングを行う。次に、電子カルテの記事の医療用語を傷病名、愁訴、身体所見、検査、治療のそれぞれに分類を行う。これらの医療用語の出現とMACEの関係を機械学習（サポートベクターマシンやディープラーニング等）の手法を用いて学習を行い、電子カルテ記事からMACEかどうかを判定する予測モデルを作成する。

（倫理面への配慮）

協力病院からデータを収集する際に、連結可能匿名化とし、個人識別情報および対応表は施設管理者の保管元、施設外に持ち出さないように厳格に管理する。

C. 研究結果

本年度（平成29年度）は、電子カルテシステム上のデータの確認と、機械学習を用いた症状記載の自動抽出に関する実験、自然言語処理を行う準備である医療用語辞書の作成、SS-MIX2データからのデータベースを構築と電子カルテ記事の抽出を行った。

（1）電子カルテシステム上のデータの確認

リアルワールドにおける電子カルテ記事の分析を行うための基盤構築にとっては、電子カルテシステムを中心とする病院情報システムにどのような情報（データ）が保存されているかの把握が必須である。そのため、現行電子カルテシステムで処理されているデータについて、その所在、データ形式について調査した。

国立循環器病研究センター電子カルテ（NEC MegaOak HR）における記述情報（SOAP、退院サマリ）の抽出を行うために、データベースのテーブル構造の把握と実際の抽出作業を実施した。その結果、文字コードの問題などがあったが、必要な記述情報を抽出できることを確認した。この際、記述情報の匿名化についても複数の課題があることがわかり、新たな課題として対応方法を検討することとした。

（2）機械学習を用いた症状記載の自動抽出に関する実験

電子カルテシステム内に蓄積された所見・報告書・サマリなどのテキスト情報から、自然言語処理および機械学習を用いて、カルテ記載内における「症状記載」について、判別・予測する方法論の検討を行った。

具体的には、臨床研究業務担当者が実際に必要とする症状記載データについて、カルテ記載情報から手動で抽出を行った。これらを用いて教師データを作成し、カルテ記載における「症状記載」と「その他の記載」についての自動判別器を作成した。自動判別器は、文章内に出現した各形態素を1次元とした線形サポートベクターマシンを用いて作成した。10分割交差検定を行い評価した結果、本判別器の感度・特異度はともに70～80%の性能を有していることがわかった。（図1）

出現形態素:13856種類

10分割交差検定の平均値

		予測値	
		正	負
実際値	正	816	1436
	負	296	7452

Accuracy (正答率) : 0.8268±0.0233
 False Negative : 63.8% (1436/2252)
 False Positive : 3.8% (296/7748)

図1. 全単語を用いた自動判定結果

(3) 自然言語処理を行う準備である医療用語辞書の作成

国立循環器病研究センターにおいて電子カルテ記事の抽出を行い、電子カルテ記事の自然言語処理を行う準備である医療用語辞書の準備を行った。平成29年度は、専門医2名、統計学者2名が国立循環器病研究センターにおいて、電子カルテデータの自然言語処理を行い、医学用語の意味体系(オントロジー)の構築とそれを利用した単語間の相関の度合い(距離等)の利用、形態素解析(名詞、助詞、動詞等の分かち書き)、係り受け解析(主語、述語等の単語間の関係)など文法の解析精度の向上を試みた。約60万行のカルテ記事を読み込み、症候の出現頻度を患者ごとに集積し、文章単位での解析が可能なため症候の出現時期、時間が同定可能であった。

さらに自然言語処理技術に関して先進的なIBMワトソンによりMajor Cardiac eventをとらえることを目的に辞書チューニングを行った。心筋梗塞レジストリMIDAS研究を中心とした約2000人の国立循環器病センター入院患者に関して、最も記述が的確と考えられる退院時サマリの記述をもとに虚血性心疾患、心不全、脳卒中、心臓死、全死亡に関してIBMワトソンエクスプローラーにより抽出を行った。死亡イベントに関しては、電子カルテ上の死亡退院により100%の把握が可能であった。初回の入院に関しては、入院契機が虚血性心疾患、心不全、脳卒中である場合もほぼ捕捉可能であった。死亡と入院契機の虚血性心疾患、心不全、不整脈項目により心臓死の確認が可能であった。辞書チューニング前はaccuracyとして65%前後であるが、チューニング後は95%以上の精度達成が可能であった。

(4) SS-MIX2 データからのデータベースを構築と電子カルテ記事の抽出

東北大学、自治医科大学、九州大学では、csv形式で出力された心電図、心臓超音波検査、心臓カテーテル検査結果を日本循環器学会標準規格

であるSEAMATに変換するためのプログラムの実装を行った。(図2)

枝振り情報・PCI座標入力モジュール

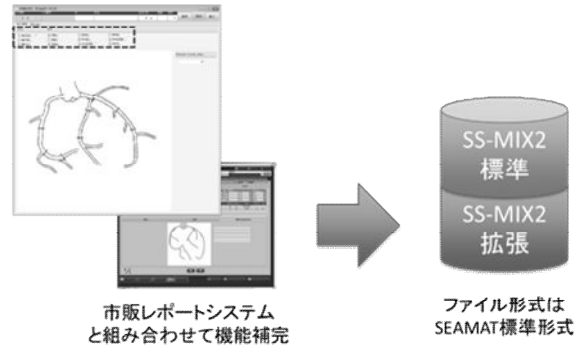


図2. 今後のCAIRS-PCIのデータの流れ

また、関連する学会との意見調整を行うため、SEAMAT研究会を発足させ、項目の見直しを行った。とりわけ、心臓超音波検査項目に対して心エコー図学会から、より実践的かつ網羅的な提案がなされ、改訂に取り組んでいる。さらに、ISO取得に向け活動の幅を広げている。東大病院の循環器系生理機能検査データ(心電図、心エコーなど)に関しては日本循環器病学会標準出力フォーマット(SEAMAT形式)への変換表を作成し、2017年11月より心エコーデータはSS-MIX2拡張ストレージへ出力が開始されている。また、その他の生理検査は2018年2月よりSS-MIX2拡張ストレージへ出力開始予定である。熊本大学では、電子カルテからは、データウェアハウスDWHに情報連携、蓄積がなされているが、このDWHから患者基本情報、病名情報、外来受診情報、入院退院情報、処方オーダ、注射オーダ、検体検査オーダ、放射線オーダ、検体検査結果、心電図数値データ、心エコー数値データ、心カテ記録、退院サマリ、経過記録に関してSS-MIX2標準ストレージ、拡張ストレージにデータ出力ならびに提供ができる状況を整えることが出来た。

D. 考察

(1) 本研究の目的である非構造化データ(テキスト情報)の自然言語処理や機械学習をするためには、対象とする所見・報告書・サマリなどのテキスト情報の所在・保管形式・データ形式などを把握する必要がある。今回の結果から、様々なシステムで作成・保管され、形式も多様なテキスト情報が、本研究で利用できる形式で抽出・収集可能か検証することができた。また、これらの結果は、他施設における情報抽出・収集においてもフィードバック可能である。そのため、多施設間で大規模なデータを収集する際には有用な知見となりうる。

(2) 本研究の最終的な目的は、電子カルテシステム内に蓄積された所見・報告書・サマリなどのテキスト情報から、自然言語処理および機械学習を用いて、Major Adverse Cardiac Event (MACE) とよばれる主要有害心血管イベントを予測するモデルを構築することである。今回行った結果から、自然言語処理を用いた機械学習が症状記載などのイベント判別・予測に有用であることが示された。

(3) 辞書チューニングの過程で抽出された構文からは、看護師、医学部生、研修医程度の精度の症候抽出は可能であり、今後登録研究における省力化、入力 of 正確性向上に有用と考えられた。

(4) 本研究結果により各施設に散在する諸検査結果の収集が可能となり、全国レベルで循環器領域における必須なデータが蓄積しうる。さらに、項目間の違いや表記ぶれ、単位の統一など、データクレンジングに必要な決まりごとを日本における循環器専門医の合意を得て行うため、大規模データを扱う上で大変重要な意義がある。また、現在医療情報分野で課題となっている SS-MIX2 拡張ストレージの充実という点でも、他の学会に先駆けて取り組んでいることは注目に値し、実際問い合わせも増えている。複数病院が参加する共同研究においては標準化した情報の連携を行い、確実な情報の収集が必要であるので今回の成果は大変意義がある。

E. 結論

本研究により、病院情報システムから、SOAP や退院サマリ、種々の検査報告書など、必要な情報を簡便に抽出できる仕組みとして、基幹システムや部門システムのデータを集約・管理できる統合DBの開発が可能となると考えられる。MACE に関連するイベントを精査し、そのイベントの判別に必要な教師データの精度の向上を行えば、機械学習手法によるより最適な予測手法が可能となる。

F. 健康危険情報 なし

G. 研究発表

1. 論文発表

- (1) 平松 治彦:医療情報システムのデータ利用における課題, Jpn Pharmacol Ther (薬理と治療), Vol.45 suppl.2, s76-s78, 2017
- (2) 平松 治彦:【改正個人情報保護法】医学研究編 国際共同研究など外国にある第三者へのデータ提供について注意すること, 医療情報学, 37(5), 253-5, 2017.
- (3) 櫻井理紗、竹村匡正、山口雅和、中井隆史、

宍戸稔聡、平松治彦、山本剛、奈良崎大士、上村幸司:ICF を用いた健康情報基盤構築のためのデータ集積手法の検討, 第 37 回医療情報学連合大会論文集, 788-789, 2017

- (4) 櫻井理紗、竹村匡正、桑直人、岡本和也、黒田知宏: 我が国における openEHR/アーキタイプを用いた診療データベースの構築可能性の検証, Mumps, vol.28, 15-23, 2017
- (5) 山田ひとみ、竹村匡正、桑田成規: 電子カルテの質向上のための診療録監査支援システムの試験的構築, Mumps, vol.28, 3-13, 2017
- (6) 山田ひとみ、竹村匡正、岡本和也、黒田知宏、桑田成規: インフォームド・コンセント記載を対象とした診療録監査システムの検討, 日本診療情報管理学会誌 29(1), 53-61, 2017

2. 学会発表

- (7) 医療情報連合大会 2017年11月22日 日本循環器学会合同シンポジウム 人工知能応用による自然言語処理の活用 電子カルテ情報のセマンティック登録と全国登録事業への将来展望
- (8) 第 82 回日本循環器学会学術集会シンポジウム 11 (2018年3月24日;大阪市) 「わが国の循環器医療提供体制の課題と展望」 The Current Status of Cardiovascular Medicine in Japan; Insights from JROAD and JROAD-DPC Database
- (9) Informatics for Health 2017(2017年4月)Poster 『Release of the Standard Export Data Format by the Japanese Circulation Society for Standardized Structured Medical Information eXchange Extended Storage』 Masaharu Nakayama.
- (10) 第 53 回日本小児循環器学会総会・学術集会 (2017年7月) 教育シンポジウム 『循環器領域におけるビッグデータ活用の道標: SS-MIX や日本循環器学会出力標準フォーマット (SEAMAT) について』 中山雅晴
- (11) 第 37 回日本医療情報学連合大会(2017年11月)共同企画シンポジウム 『循環器領域におけるビッグデータ活用の現在』 中山雅晴
- (12) 第 37 回日本医療情報学連合大会(2017年

11 月) 一般口演 『SS-MIX2 拡張ストレージの充実に向けた取り組み - 日本循環器学会出力標準フォーマット(SEAMAT)について -』中山雅晴、竹花一哉、興梠貴英、IHE-J 循環器

- (1 3) 日本循環器学会総会(平成 30 年 3 月 25 日、大阪)「臨床効果データベース事業・ImPACT 研究におけるデータ収集の現状」
- (1 4) 平松治彦:シンポジウム 1「pragmatic clinical trial への誘い」医療情報システムのデータ利用における課題, 日本臨床試験学会第 8 回学術総会
- (1 5) 櫻井理紗、竹村匡正、山口雅和、松本佳久、本谷崇之、今津貴史、上村幸司、平松治彦、山本剛、奈良崎大士、宍戸稔聡: ICF を用いた個人健康管理システムの構築, 第 44 回日本 M テクノロジー学会大会
- (1 6) 櫻井理紗、竹村匡正、山口雅和、中井隆史、宍戸稔聡、平松治彦、山本剛、奈良崎大士、上村幸司: ICF を用いた健康情報基盤構築のためのデータ集積手法の検討, 第 37 回医療情報学連合大会

H. 知的財産権の出願・登録状況

- 1. 特許取得 なし
- 2. 実用新案登録 なし
- 3. その他 なし

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

自然言語処理による MACE イベント捕捉の検討に関する研究

研究分担者 西村 邦宏 国立循環器病研究センター・予防医学・疫学情報部・部長
中村 文明 国立循環器病研究センター・循環器病統合センター・室長
竹上 未紗 国立循環器病研究センター・予防医学・疫学情報部・室長
安田 聡 国立循環器病研究センター・副院長・心臓血管内科部門長
宮本 恵宏 国立循環器病研究センター・循環器病統合センター・センター長

研究要旨

これまでの研究において、電子カルテシステム内に蓄積された所見・報告書・サマリなどのテキスト情報から、自然言語処理および機械学習を用いて、カルテ記載内における「症状記載」について、判別・予測する方法論の検討を行った。臨床研究業務担当者が実際に必要とする症状記載データについて、カルテ記載情報から手動で抽出を行った。これらを用いて教師データを作成し、カルテ記載における「症状記載」と「その他の記載」についての自動判別器を作成した。こちらは感度・特異度はともに70~80%の性能を有していることがわかった。

この結果をもとにさらに自然言語処理技術に関して先進的な IBM ワトソンにより Major Cardiac event をとらえることを目的に辞書チューニングを行った。心筋梗塞レジストリ MIDAS 研究を中心とした約2000人の国立循環器病センター入院患者に関して、最も記述が的確と考えられる退院時サマリの記述をもとに虚血性心疾患、心不全、脳卒中、心臓死、全死亡に関して IBM ワトソンエクスプローラーにより抽出を行った。死亡イベントに関しては、電子カルテ上の死亡退院により100%の把握が可能であった。初回の入院に関しては、入院契機が虚血性心疾患、心不全、脳卒中である場合もほぼ捕捉可能であった。死亡と入院契機の虚血性心疾患、心不全、不整脈項目により心臓死の確認が可能であった。辞書チューニング前は accuracy として65%前後であるが、チューニング後は95%以上の精度達成が可能であった。

入院時における血性心疾患、心不全、脳卒中、心臓死、全死亡に関して IBM ワトソンエクスプローラーによる自然言語処理でほぼ正確に MACE の捕捉が可能であった。

A．研究目的

前年度までの研究において、電子カルテシステム内に蓄積された所見・報告書・サマリなどのテキスト情報から、自然言語処理および機械学習を用いて、カルテ記載内における「症状記載」について、判別・予測する方法論の検討を行った。臨床研究業務担当者が実際に必要とする症状記載データについて、カルテ記載情報から手動で抽出を行った。これらを用いて教師データを作成し、カルテ記載における「症状記載」と「その他の記載」についての自動判別器を作成した。こちらは感度・特異度はともに70~80%の性能を有していることがわかった。

この結果をもとにさらに自然言語処理技術に関して先進的な IBM ワトソンにより Major Cardiac event をとらえることを目的に辞書チューニングを行った。

B．研究方法

心筋梗塞レジストリ MIDAS 研究を中心とした約2000人の国立循環器病センター入院患者に関して、最も記述が的確と考えられる退院時サマリの記述をもとに虚血性心疾患、心不全、脳卒中、心臓死、全死亡に関して IBM ワトソンエクスプローラーにより抽出を行った

(倫理面への配慮)

- 6 人を対象とする医学系研究に関する倫理指針を遵守し研究を遂行する。

C. 研究結果

死亡イベントに関しては、電子カルテ上の死亡退院により 100%の把握が可能であった。初回の入院に関しては、入院契機が虚血性心疾患、心不全、脳卒中である場合もほぼ捕捉可能であった。死亡と入院契機の虚血性心疾患、心不全、不整脈項目により心臓死の確認が可能であった。辞書チューニング前は accuracy として 65%前後であるが、チューニング後は 95%以上の精度達成が可能であった。

D. 考察

入院契機ではなく例えば予定 PCI として紹介され、既往歴としてイベントとなる病名が記述されている場合は、日付けが何年何月など明記されず入院時から去年など曖昧さを残す記述であることがあり、これら相対日付けに関しては今後の精度向上が必要と考えられた

E. 結論

入院時における血性心疾患、心不全、脳卒中、心臓死、全死亡に関して IBM ワトソンエクスプロラーによる自然言語処理でほぼ正確に MACE の捕捉が可能であった。

G. 研究発表

1. 論文発表 なし
2. 学会発表
 - 1) 医療情報連合大会 2017年11月22日
日本循環器学会合同シンポジウム
人工知能応用による自然言語処理の活用
電子カルテ情報のセマンティック登録と全国登録事業への将来展望

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

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

臨床効果データベース事業におけるデータ出力方法の検討に関する研究

研究分担者 安田 聡 国立循環器病研究センター・副院長・心臓血管内科部門長
斉藤 能彦 奈良県立医科大学・循環器内科・教授

研究要旨

日本循環器学会の事業である「臨床効果データベース」では、循環器領域で特にその重要性が指摘されている疾患を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである。心不全は再発を繰り返しながら心機能低下・死亡へと至る。心不全の増悪による再入院をいかに起こさないようにするかが、患者の生活の質を維持するために必要であり、その実態を DPC データ(JROAD-DPC ; 2012-2014 年度 3 年間)をもとに解析したので報告する。

A . 研究目的

日本循環器学会の事業である「臨床効果データベース」では、循環器領域で特にその重要性が指摘されている疾患(心不全など)を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである。心不全は再発を繰り返しながら心機能低下・死亡へと至る。心不全の増悪による再入院をいかに起こさないようにするかが、患者の生活の質を維持するために必要であり、その実態を DPC データ(JROAD-DPC)をもとに解析した。

B . 研究方法

循環器疾患診療実態調査(日本循環器学会主導、全国循環器専門医研修施設・研修関連施設 1335 施設)の枠組みを用いて、診断、短期予後、年齢、性別、合併症、重症度、使用薬剤などが含まれる DPC (Diagnosis Procedure Combination ; 診断群分類包括評価)情報を収集した「DPC データを用いた心疾患における医療の質に関する事業」(JROAD-DPC)を行った。

(倫理面への配慮)

人を対象とする医学系研究に関する倫理指針を遵守し研究を遂行する。

C . 研究結果

2012-2014 年度の 3 年間 DPC データを解析、心不全患者延べ 216,157 例(M; 75 ± 13, F; 81 ± 12 歳)を抽出した。本邦では 3 年間で 20%が複数回の入院(再入院)をしているという実態が明らかになった。

D . 考察

これらの結果は既報のレジストリ研究 Chronic Heart Failure Analysis and Registry in the Tohoku District (CHART)-2 研究(N=4,682)での 3 年間観察中 17%の再入院という結果も概ね一致するものであった(Circ J. 2015;79:2396-407)。

E . 結論

「臨床効果データベース」において、縦断的データベースとして心不全をモデル疾患とした情報収集体制を整備する。得られたデータは DPC データ、レジストリデータと比較検証していく。

G . 研究発表

1. 論文発表 なし
2. 学会発表
(1) 第 82 回日本循環器学会学術集会シンポジウム 11(2018 年 3 月 24 日;大阪市)
「わが国の循環器医療提供体制の課題と展望」
The Current Status of Cardiovascular Medicine in Japan; Insights from JROAD and JROAD-DPC Database

H . 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

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

データ転送プログラムによるデータ収集に関する研究

研究分担者 中山 雅晴 東北大学大学院・医学系研究科・教授

研究要旨

日本循環器学会で標準出力フォーマット(Standard Export data for MAT: SEAMAT)を定めることにより、厚生労働省標準保存形式であるSS-MIX2の拡張ストレージに循環器特有の検査結果を転送し、データを2次活用するための基盤を整備した。本年度はさらなる項目整備のため、SEAMAT研究会を設立、日本循環器学会の他、日本医療情報学会、日本心不全学会、日本不整脈心電学会、心エコー図学会、日本心血管インターベンション治療学会、日本核医学会が参加し、改訂作業を行った。また、SS-MIX2 agentを各施設のサーバーに設置することにより、SS-MIX2経由の循環器データが処方や検査値とともに抽出されることを確かめた。今後、対応施設の拡大とデータの検証作業を行う。

A. 研究目的

MID-NETの本格運用をはじめ、厚生労働省電子的診療情報交換推進事業に基づいたStandardized Structured Medical Information eXchange version 2 (SS-MIX2)ストレージ形式を用いたデータ活用が始まっている。しかしながら、利用されているデータのほとんどはSS-MIX2の標準化ストレージであり、病名や検査値、処方などが該当する。我々は、循環器疾患の診断・治療に特有の検査および治療データの活用をもとに、SS-MIX2拡張ストレージへデータを出力するための標準フォーマットStandard Export data for MAT (SEAMAT)を日本循環器学会やIHEの協力の下2015年度に定めた。実際にSEAMAT仕様を搭載した生理検査システムも販売され始めているが、臨床医や医療情報専門家として、これらの項目の妥当性や各施設から出力されるデータの抽出作業の容易さについて検討を行う必要がある。

B. 研究方法

本研究対象施設である国立循環器病センター、東京大学、東北大学、九州大学、自治医大における生理検査システムや心臓カテテル検査レポートシステムからSEAMATによるデータ出力を行う。実現するための手段として、SEAMAT対応のシステムからの出力、各システムからデータをcsvやxml形式で出力し、SEAMAT形式に変換してSS-MIX2に出力することが挙げられる。東北大、九州大、自治医大、東大でそれぞれの施設からデータ出力を行う。またSS-MIX2からSS-MIX2agentという製品を用いてデータ抽出も試みる。

(倫理面への配慮)

実名情報をストレージで扱う範囲ではそれぞれの施設の院内ルールに従う。ストレージからの情報共有に関しては、本研究に関するそれぞれの施設の倫理委員会の裁定に遵守し、とりわけ個人情報の保護の観点からその取扱いには十分に配慮を行う。

C. 研究結果

csvファイルからのデータを転送するプログラムとともにxmlからの対応も可能とした。SEAMATの項目に関しては、心エコー図学会より改訂の申し入れがあり、現在改訂作業中である。

平成29年度、SS-MIX2 agentにより、各施設から心電図や心臓超音波、心臓カテテル検査結果などとともに、病名、処方、採血結果なども抽出されることを確認した。

D. 考察

ビッグデータ時代とはいえ、必要なデータは各施設の中にあり、実はそれらが十分に活用されていない。本事業は、循環器疾患を研究することにおいて必須のデータを収集、活用するための必要基盤である。

今後、リアルワールドデータとして、電子カルテすなわちSS-MIX2ストレージ内の情報を用いることが主流となっていくことが予想されるが、MID-NETを始めとして、標準化ストレージのみのデータ収集ではすぐに限界に直面するであろう、また画像解析においても、患者基本および診療情報の蓄積が必須であることも早晚理解されると思う。我々の目指すところは現場医師の手入力によるデータ転送で無く、数多あるデータのシステムチツ

クな活用である。

今後も日本循環器学会のIT/DB委員会を中心に各関連学会のコンセンサスを取りながら実装を進めていき、国際間でのデータ交換規約へと発展させたい。

E．結論

各循環器関連学会、医学情報専門家、ベンダーの協力の下、循環器領域のデータを効率的に収集、2次活用するためにSEAMATに準拠したデータ収集の基盤を整備している。今後他施設へ展開させていく。

G．研究発表

1. 論文発表 なし
2. 学会発表

(1) Informatics for Health 2017(2017年4月)Poster

『Release of the Standard Export Data Format by the Japanese Circulation Society for Standardized Structured Medical Information eXchange Extended Storage』 Masaharu Nakayama.

(2) 第53回日本小児循環器学会総会・学術集会(2017年7月)教育シンポジウム

『循環器領域におけるビッグデータ活用の道標：SS-MIXや日本循環器学会出力標準フォーマット(SEAMAT)について』中山雅晴

(3) 第37回日本医療情報学連合大会(2017年11月)共同企画シンポジウム

『循環器領域におけるビッグデータ活用の現在』中山雅晴

(4) 第37回日本医療情報学連合大会(2017年11月)一般口演

『SS-MIX2拡張ストレージの充実に向けた取り組み - 日本循環器学会出力標準フォーマット(SEAMAT)について -』中山雅晴、竹花一哉、興梠貴英、IHE-J循環器

H．知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

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

SS-MIX2データ抽出システム実装とテスト及びアウトカムテンプレート仕様検討

研究分担者 的場 哲哉 九州大学病院・循環器内科・講師

研究要旨

本研究は電子カルテSS-MIX2ストレージに格納された多モダリティの臨床データを一括抽出するデータ登録システムの多施設における実装を目的とし、MCDRSを応用した冠動脈疾患患者の検体検査、処方、心臓カテーテル検査レポート、冠動脈インターベンションレポートデータの半自動登録システムを実装し、さらに心血管病アウトカムを補足する症例登録フォームテンプレートを開発している。

A. 研究目的

電子カルテSS-MIX2ストレージに格納された多モダリティの臨床データを一括抽出するデータ登録システムの実装、および心血管病アウトカムを補足する症例登録フォームテンプレートシステムの開発

B. 研究方法

1. MCDRSを応用した冠動脈疾患患者の多モダリティ臨床データ(検体検査、処方、心臓カテーテル検査レポート)の半自動登録システムを実装する。
2. 心血管病アウトカムを補足する症例登録フォームテンプレートで登録するデータ項目を選定し、SS-MIX2へのデータ登録仕様を決定し、電子カルテシステムのテンプレートシステムとして実装する。

(倫理面への配慮)

研究施設において臨床研究倫理審査を受けた。登録する症例データは匿名化される。

C. 研究結果

1. MCDRSを応用した冠動脈疾患患者の多モダリティ臨床データ(検体検査、処方、心臓カテーテル検査レポート)の半自動登録システムを実装し、取得されたデータと原本の同一性を検討した(論文投稿中)。
2. 心血管病アウトカムを補足する症例登録フォームテンプレートで登録するデータ項目を選定し、SS-MIX2へのデータ登録仕様を決定した。電子カルテシステムのテンプレートシステムとして実装している。

D. 考察

冠動脈疾患患者の多モダリティ臨床データ(検体検査、処方、心臓カテーテル検査レポート)の単一システムによる一括収集は、現在市場にあるシステムでは実現されておらず、重要な進歩である。登録

型臨床研究の質の向上、多施設共同研究の推進に寄与することが期待される。

E. 結論

平成30年度はデータ収集システムの開発に進捗が得られた。また、心血管病アウトカム登録システムを開発中である。

G. 研究発表

1. 論文発表 なし
2. 学会発表
日本循環器学会総会(平成30年3月25日、大阪)
「臨床効果データベース事業・ImPACT 研究におけるデータ収集の現状」

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

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

臨床効果データベース事業の学会支援

研究分担者 小室一成 東京大学大学院医学系研究科・循環器内科・教授

研究要旨

臨床効果データベース事業の推進の学会支援を行っている。SS-MIX2 対応にむけた生理機能検査データの出力標準フォーマット(SEAMAT)を策定し日本循環器学会標準フォーマットとしてその普及と推進を学会として支援した。東大病院では、生理機能検査の SEAMAT 対応変換表を作成し SS-MIX2 拡張ストレージへの蓄積が開始されたところである。現在、蓄積データの検証を行うところである。臨床効果データベースの推進と拡充と並行して、研究用データベース作成へのデータ提供の準備を行っている。

A. 研究目的

電子カルテの記事情報から自然言語処理を活用して自動的に Major Adverse Cardiac Event (MACE)とよばれる主要有害心血管イベントを判断するシステムを開発し、電子情報による医療ビッグデータ自動分析を行うためのシステムを構築する。

B. 研究方法

心臓カテテル検査、生理機能検査を SS-MIX2 拡張ストレージに蓄積を行い、臨床データ登録システム MCDRS を用いて研究用データベースを作成する。東大病院診療データを国立循環器病研究センターへ連結可能匿名化を行い提供する。

(倫理面への配慮)

東京大学医学部附属病院のカルテ情報の利用にあたって、東京大学医学部倫理委員会から承認を受けた。データを収集する際に個人情報 は削除して連結可能匿名化とし、個人識別情報 および対応表を施設外に持ち出さないように厳格に管理する。

C. 研究結果

SS-MIX2 対応にむけた生理機能検査のデータ標準化に関して、中山雅晴先生が中心となり日本循環器学会データ出力標準フォーマット (SEAMAT) を策定した。日本循環器学会標準フォーマットとしてその普及と推進を学会として支援を行った。また循環器関連学会による SEAMAT 研究会にも日本循環器学会として参加している。東大病院では医療情報部教授大江和彦先生の協力のもと、生理機能検査の保存形式である

CSV 形式から SEAMAT 形式への変換表を作成し、SS-MIX2 拡張ストレージへのデータ転送準備がほぼ完了したところである。東大病院では平成29年11月より生理機能検査の SS-MIX2 拡張ストレージへの蓄積が開始された。SS-MIX2 拡張ストレージ上の蓄積データの検証を行い、過去のデータの転送準備を進める予定である。SS-MIX2 拡張ストレージの拡充に並行して、研究用データベース作成のためのデータ提供の準備を行っている。データは匿名化など、国立循環器病研究センターへの提供準備を行う。

D. 考察

SS-MIX2 および MCDRS を用いたデータベース事業として糖尿病の J-DREAMS 研究、腎臓病の J-CKD-DB 研究が先行している。循環器領域特有の多様な診断モダリティの SS-MIX2 対応が進むことで、循環器領域での大規模データベース事業が進展することが期待される。

E. 結論

本年度は SEAMAT 変換表を作成し、生理機能検査データの SS-MIX2 への払い出しが始まったところである。MCDRS を用いたデータ抽出を行い、問題点の洗い出しを行っているところである。

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成 29 年度厚生労働省科学研究補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業))
分担研究報告書

自然言語解析による診断名判断システムの開発

研究分担者 穴戸 稔聡 国立循環器病研究センター・研究推進支援部・部長
平松 治彦 国立循環器病研究センター・情報統括部・部長
上村 幸司 国立循環器病研究センター・研究推進支援部・室長
竹村 匡正 兵庫県立大学大学院・応用情報科学研究科・教授

研究要旨

診断名の判断には、電子カルテの蓄積データ、特に、自然文で記載されるカルテ記事が重要である。また、カルテ記事の分析においては、関連する情報を利用することで分析精度の向上や効率化が図れると考えている。そこで、電子カルテシステム内のデータの所在や形式について調査し、文字コードの問題があったが、必要な記述情報を抽出することができた。今後、必要な情報を簡便に抽出できる仕組みとして、基幹システムや部門システムのデータを集約・管理できる統合 DB の開発を行う予定である。また、データ抽出の際、適切な匿名化、暗号化を適用するための仕組みと匿名化のための辞書の整備も検討する必要があることがわかった。

次に、電子カルテシステム内に蓄積された所見等の記述情報から、自然言語処理および機械学習を用いて、「症状記載」と「その他の記載」に分類する自動判別器を作成した。得られた自動判別器は、正答率 80%以上、感度・特異度ともに 70~80%の性能を有していた。今後は、より多くの高精度な教師データの作成と、いくつかの機械学習手法による比較評価および最適な手法の検討を行っていく。

A. 研究目的

本研究の目的は、電子カルテシステム内に蓄積された所見・報告書・サマリなどの記述情報から、自然言語処理および機械学習を用いて診断名を推測、判断できる枠組みを構築することにある。

しかし、対象とする記述情報は、様々なシステムで作成・保管され、形式も多様であるため、本研究で利用できる形式で抽出・収集可能か確認が必要である。

また、カルテに記載されている内容は既往歴・主訴・現病歴・症状・検査結果・評価・治療方法(処置・処方・手術など)・経過など様々である。その中には、病名・治療方法などの医学用語、日付・検査値などの数値情報、人名などの個人情報等、種々の情報が混在しており、自然言語処理による形態素解析やそれに伴う機械学習の精度の向上には多くの問題が想定される。

そこで、今回、第 1 段階として、電子カルテシステムに保管されている記述情報について所在や形式を確認し、抽出されたデータについて検証を行い、自然言語処理や機械学習を

行うことが可能か評価を行った。

次に、電子カルテシステムに記載された記述情報に対する自然言語処理および機械学習の有用性を評価するために、対象を電子カルテシステムの SOAP 記載のうち S と O の記述情報に絞り、有害事象などの把握につながる「症状記載」について判別・予測できるか検証を行った。

本稿では、以上の 2 点について報告する。

B. 研究方法

1. 電子カルテシステム内のデータ確認

当センター電子カルテシステム(NEC MegaOak HR)における記述情報(SOAP, 退院サマリなど)の抽出を行うためにデータベースのテーブル構造の調査を行った。次に、調査結果に基づきデータの抽出を行い、抽出された記述情報の内容を確認した。

2. 機械学習を用いた症状記載の自動抽出に関する検討

臨床研究業務担当者(製薬企業薬剤師)1名が、10000 件の電子カルテ記述情報(S と O)

を対象に、有害事象に関連する「症状記載」データを手動で抽出し、「症状記載」の有無を判別し教師データを作成した。

次に、上述の教師データに対して自然言語処理を用いて形態素解析を行い、カルテ記述情報の形態素（単語）を抽出した。形態素解析は、医療用語辞書（30万語）を用いる場合と用いない場合で行った。

最後に、文章内に出現した各形態素（単語）を1次元とした線形サポートベクターマシンを用いて学習させ、カルテ記述における「症状記載」と「その他の記載」の自動判別器を作成した。

作成した自動判別機の性能は、10分割交差検定を用いて評価した。

（倫理面への配慮）

本研究において診療データを利用する際には、国立循環器病研究センターなど参加施設の倫理委員会の承認を得てその指示に従う。情報収集協力病院からデータを収集する際には、個人情報削除して連結可能匿名化とし、個人識別情報および対応表を施設外に持ち出さないように厳格に管理する。

C. 研究結果

1. 電子カルテシステム内のデータ確認

所見・報告書・サマリなどの記述情報は、テキストデータとして抽出することができた。内容を精査した結果、一部、文字コードの問題による文字化けなどが発生したが、適切に文字コード変換することで、必要な記述情報を抽出することができた。

2. 機械学習を用いた症状記載の自動抽出に関する検討

エキスパート1名が手動で行った「症状記載」の有無の判別結果は、S情報：有 964件 無 3672件、O情報：有 1285件 無 4076件だった。

自然言語処理における形態素解析の結果は、医療用語辞書を用いた場合、形態素（単語）の数は3480種類、辞書を用いない場合、13856種類だった。

辞書を用いた場合と用いなかった場合の形態素（単語）を学習して作成された2種類の自動判別器を、10分割交差検定で評価した。その結果、辞書を用いなかった場合のほうが多少性能はよかったが、双方とも正答率は80%以上あり、感度・特異度ともに70~80%

の性能を有していた。

D. 考察

1. 電子カルテシステム内のデータ確認

本研究の目的である非構造化データ（記述情報）の自然言語処理や機械学習をするためには、対象とする所見・報告書・サマリなどの記述情報の所在・保管形式・データ形式などを把握する必要がある。今回の結果から、様々なシステムで作成・保管され、形式も多様な記述情報が、本研究で利用できる形式で抽出・収集可能か検証することができた。

また、抽出した記述情報の中には、本人・家族・医療スタッフ等の氏名や、人物名が含まれた病名（例：橋本病）などが混在していた。しかし、人名辞書などを用いた単純なマスキングでは、適切に匿名化することはできなかったため、辞書のチューニングや手動で匿名化を行う必要があった。今後、最適な匿名化手法についての検討も必要である。

以上の結果は、他施設における情報抽出・収集においてもフィードバック可能である。そのため、多施設間で大規模なデータを収集する際には有用な知見となりうる。

2. 機械学習を用いた症状記載の自動抽出に関する検討

今回行った結果から、自然言語処理を用いた機械学習が「症状記載」などのイベント判別・予測に有用であることが示せた。

ただし、教師データはエキスパート1名により手動で作成されたため、主観によるバイアスが混入している可能性がある。したがって、学習の精度向上のためには、複数名による教師データの精査が必要である。

また、学習効果は学習データの数にも関連するため、質の良い学習データを可能な限り増やす必要があるが、学習のための演算時間とのトレードオフでもあるため、適当な数を評価する必要もある。

形態素解析の結果、辞書有のほうが抽出単語数は少なかった。これは、医療用語が細切れにならず適切に抽出されている可能性を示唆しているが、抽出単語の精査を行った上で、辞書のチューニングを行う必要がある。

自動判別機の性能は、辞書有と無であまり差はなかった。辞書有の場合、抽出された単語数が辞書無の4分の1以下だったため、学習時間を考えると、辞書有で学習データを作成する方が有用であると考えられる。

さらに、機械学習のアルゴリズムはSVM以外にも何十種類もあり、学習方法も異なる。最も優れた手法や何にでも使える手法というものはないため、適切なアルゴリズムを探すには試行錯誤に頼らざるを得ない部分がある。ただし、アルゴリズムの選択は、扱うデータのサイズや種類、データから導き出したい見解、その見解の活用方法によって決まってくる部分もあるため、先行事例を参照しながら、より最適なアルゴリズムの検討を行っていく必要がある。

E. 結論

自然文で記載される電子カルテ記事の分析を行うために、電子カルテシステム内のデータの所在や形式について調査した。文字コードの問題などがあったが、必要な記述情報を抽出できることが確認できた。また、データ抽出の際、適切な匿名化、暗号化などができる仕組みも検討する必要があることがわかった。今後、病院情報システムから、SOAPや退院サマリ、種々の検査報告書など、必要な情報を簡便に抽出できる仕組みとして、基幹システムや部門システムのデータを集約・管理できる統合DBの開発を行う予定である。

電子カルテシステム内の記述情報(SとO)から、自然言語処理および機械学習を用いて、「症状記載」と「その他の記載」の自動判別器を作成した。得られた自動判別器は、正答率80%以上、感度・特異度ともに70~80%の性能を有していた。今後は、MACEに関連するイベントを精査し、そのイベントの判別に必要な教師データの作成や質の高い学習データの作成に必要な医療用語辞書のチューニングを行う。また、SVMだけではなく、ディープラーニングなどいくつかの機械学習手法による比較評価および最適な手法の検討を行う。

G. 研究発表

1. 論文発表

- 1) 平松 治彦:医療情報システムでのデータ利用における課題, Jpn Pharmacol Ther (薬理と治療), Vol.45 suppl.2, s76-s78, 2017
- 2) 平松 治彦:【改正個人情報保護法】医学研究編 国際共同研究など外国に

ある第三者へのデータ提供について注意すること, 医療情報学, 37(5), 253-5, 2017.

- 3) 櫻井理紗, 竹村匡正, 山口雅和, 中井隆史, 宍戸稔聡, 平松治彦, 山本剛, 奈良崎大士, 上村幸司: ICFを用いた健康情報基盤構築のためのデータ集積手法の検討, 第37回医療情報学連合大会論文集, 788-789, 2017
- 4) 櫻井理紗, 竹村匡正, 糸直人, 岡本和也, 黒田知宏: 我が国におけるopenEHR/アーキタイプを用いた診療データベースの構築可能性の検証, Mumps, vol.28, 15-23, 2017
- 5) 山田ひとみ, 竹村匡正, 桑田成規: 電子カルテの質向上のための診療録監査支援システムの試験的構築, Mumps, vol.28, 3-13, 2017
- 6) 山田ひとみ, 竹村匡正, 岡本和也, 黒田知宏, 桑田成規: インフォームド・コンセント記載を対象とした診療録監査システムの検討, 日本診療情報管理学会誌 29(1), 53-61, 2017

2. 学会発表

- 1) 平松治彦: シンポジウム1「pragmatic clinical trial への誘い」医療情報システムのデータ利用における課題, 日本臨床試験学会第8回学術総会
- 2) 櫻井理紗, 竹村匡正, 山口雅和, 松本佳久, 本谷崇之, 今津貴史, 上村幸司, 平松治彦, 山本剛, 奈良崎大士, 宍戸稔聡: ICFを用いた個人健康管理システムの構築, 第44回日本Mテクノロジー学会大会
- 3) 櫻井理紗, 竹村匡正, 山口雅和, 中井隆史, 宍戸稔聡, 平松治彦, 山本剛, 奈良崎大士, 上村幸司: ICFを用いた健康情報基盤構築のためのデータ集積手法の検討, 第37回医療情報学連合大会

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成29年度厚生労働省科学研究補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業))
分担研究報告書

SS-MIX2を利用した臨床効果データベースのデータ収集に関する研究

研究分担者 辻田 賢一 熊本大学大学院・生命科学研究所・教授
宇宿 功市郎 熊本大学医学部附属病院・医療情報経営企画部・教授

研究要旨

本事業の臨床効果データベースに必要な項目が、熊本大学医学部附属病院に構築したDWHに存在し、SS-MIX2標準及び拡張データとして出力可能か確認した。

A. 研究目的

熊本大学医学部附属病院では平成29年1月に病院内情報システムを更新した際に電子カルテの情報をDWHに蓄積しそこからSS-MIX2へ変換出力する環境を構築した。これは本事業の昨年度研究に基づくものである。本年度は、この構築したDWHに本事業で収集対象となる診療情報が存在しているかどうかの調査を行い、SS-MIX2データとして外部提供を行えるかどうかの検討を行うこととした。

B. 研究方法

平成29年1月に病院内情報システムを更新した際に構築したDWHに蓄積された情報の内容と、本事業で必要な情報を項目レベルで突合せてお互いの整合性が取れているかどうかの調査を実施した。この調査結果から本事業で必要な項目のうちDWHに不足している項目に関してデータの所在確認を行った。

(倫理面への配慮)

本事業は臨床情報の外部提供が必要なため情報連携についての院内掲示を行うとともに、本事業について倫理委員会へ研究の申請を行い、承認を得た。

C. 研究結果

DWHからは、SS-MIX2標準データに該当する、患者基本情報、病名情報、外来受診情報、入退院情報、処方オーダ、注射オーダ、検体検査オーダ、放射線オーダ、検体検査結果と、SS-MIX2拡張データに該当する、経過記録、退院サマリについて、本事業に必要な項目と合致することが確認できた。また、心電図数値データ、心エコー数値データ、心カテ記録については、循環器分野のSS-MIX2拡張データを出力するための標準フォーマットStandard Export data for MAT (SEAMAT)との整合性を確認した結果、心電図では100%(メーカー依

存情報を除く)、心エコーでは54.6%、心カテ記録では71.6%の合致を認めた。不足項目のうち心エコーは、検査モダリティ内に存在しDWHに送信されていなかった。心カテは、熊大のカテレポートシステムの入力項目に存在しなかった。

D. 考察

今回の研究において、本事業に必要な非構造化データ(テキスト情報)がSS-MIX2拡張データとして出力可能な状況にあることが確認できた。構造化データについてもSS-MIX2標準及び拡張データとして大部分が出力可能な状況にあることが確認できた。不足項目についてはDWHへ蓄積できるように今後各システムベンダーと調整を行う予定である。また、本事業に関して外部へのデータ提供を含む臨床研究の申請を行い、熊本大学の倫理委員会の承認を得た。本研究事業を通じ、当施設の問題点を洗い出しながらデータベースの構築やデータ抽出、出力の標準化を展開していく必要がある。

E. 結論

本事業に必要な項目の大部分が既にDWHに蓄積され、SS-MIX2標準及び拡張データとして出力が可能であることを確認できた。

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
なし							

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
平松 治彦	医療情報システムのデータ利用における課題	Jpn Pharmacol Ther (薬理と治療)	Vol.45 suppl.2	s76-s78	2017
平松 治彦	[改正個人情報保護法] 医学研究編 国際共同研究など外国にある第三者へのデータ提供について注意すること.	医療情報学	37(5)	253-255	2017
櫻井理紗、竹村匡正、山口雅和、中井隆史、宍戸稔聡、平松治彦、山本剛、奈良崎大士、上村幸司	ICFを用いた健康情報基盤構築のためのデータ集積手法の検討	第37回医療情報学連合大会論文集		788-789	2017
櫻井理紗、竹村匡正、糸直人、岡本和也、黒田知宏	我が国におけるopenEHR/アーキタイプを用いた診療データベースの構築可能性の検証	Mumps	vol.28	15-23	2017
山田ひとみ、竹村匡正、桑田成規	電子カルテの質向上のための診療録監査支援システムの試験的構築	Mumps	vol.28	3-13,	2017
山田ひとみ、竹村匡正、岡本和也、黒田知宏、桑田成規	インフォームド・コンセント記載を対象とした診療録監査システムの検討	日本診療情報管理学会誌	29(1)	53-61	2017