

平成 29 年度厚生労働科学研究補助金

(臨床研究等 ICT 基盤構築・人工知能研究事業) 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

堀口 裕正 国立病院機構本部総合研究センター 診療情報分析部 副部長

岡田 千春 国立病院機構本部 企画役

狩野 芳伸 静岡大学情報学部行動情報学科 准教授

森田 瑞樹 岡山大学大学院医歯薬学総合研究科 准教授

奥村 貴史 国立保健医療科学院研究情報支援研究センター 特命上席主任研究官

研究要旨

初年度においては、我々が日本語における医療用自然言語処理の研究コミュニティを形成し研究に取り組んで来た標準化技術を実カルテへと適用することで、カルテからの情報抽出の自動化に向けた予備的な検証を行うことを計画した。

研究代表者堀口及び分担研究者岡田は、国立病院機構本部との調整を中心とした基盤構築を行った。まず、NDCA データの研究利用に向け、倫理審査申請に加えて、内部規定にて定められている内部委員会の調整を図った。また、閲覧・解析に特化した自然言語処理用の研究基盤の構築を行った。研究基盤の概念図は図に示したとおりで、セキュリティを維持しつつ、空間的制約をなるべく少なく研究が進められるようなものになっている。

研究分担奥村は、臨床的なニーズを自然言語処理における個別技術へと橋渡しする役割を担った。具体的には、臨床医側より退院サマリの自動生成に求められる要件定義を進めるとともに、先行研究の整理を行い今後の研究アプローチの策定を行った。さらに、今後の研究に役立てられる入院カルテ・退院サマリの高品質な個人情報を含まない模擬のデータセットを構築した。

研究分担狩野・森田は、上記の実データ・テスト用データ双方を活用し、自然言語処理の医療テキストへの適用を進めた。狩野は、時系列で蓄積してく入院カルテデータを対象として、既存の自然言語処理ツールによる処理性能と今後の改良に向けた課題抽出を図った。森田は、医師が要約した退院サマリデータを対象として、医師の記載する退院サマリの定量的・定性的な特徴の把握を図った。この知見は、今後、入院カルテの自動要約技術

の研究に際した精度管理に役立てられる。

今年度の研究においては、NDCA データの利用に際した倫理審査の関係で、実カルテを対象とした解析は 3 月以降に限られる制約が生じた。しかしながら、高品質なテストデータの確保が可能であったことから、研究のコアである医療用自然言語処理部分の検討はほぼ計画通りに進めることが出来た。電子カルテの自由記載部分を自動解析する多施設構成での大規模データを対象とした研究としては、本邦初の試みとなる。

医療用情報システムの研究開発においては、医療現場に直接の恩恵が及ばないゴールが設定されることで、研究開発が現場のニーズから乖離するとともに、継続した開発投資に繋がらない悪循環が往々にして生じてきた。本研究提案は、医療現場における負担軽減策として期待が大きい退院サマリの自動要約技術の開発を目指す。これにより、紹介状の自動作成技術等、電子カルテの自動解析技術に関連する継続的な研究開発投資の実現が期待される。この研究開発サイクルを確立することにより、要素技術である電子カルテ上の記載から自動情報抽出における継続的な精度向上が期待される。

こうして確立する医療用自然言語処理技術は、大量の電子カルテからの効率的な情報抽出を実現し、健康医療政策に資する統計データの収集コストを劇的に低廉化することが期待される。とりわけ、様々な傷病や治療に関して、既存の DPC やレセプトには表れてこない深遠な実態を明らかとし、医療の質向上・均てん化・各種医療技術の臨床開発に必要なエビデンスを生み出していくことが期待される。

また、国立病院機構の有する広域電子カルテ網は、各病院が独自に調達した電子カルテベンダー主要 6 社を網羅している。本研究によって、大口顧客としての交渉力を背景としたこれら主要ベンダーへの研究開発成果の技術移転が期待される。これらは、医療の情報化を進める厚生労働行政にとって、新たな政策手段の実現をもたらす

加えて、本研究で作成した入院カルテ・退院サマリの高品質な個人情報を含まない模擬のデータセットは、研究成果として公表することにしており、電子カルテの現物を持たない情報系研究者が、容易に本領域の発展に貢献できる環境を提供出来る点も大きな成果になるものと考えている。

A. 目的

本研究は、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい退院サマリの自動生成技術の実現を目標と定める。電子カルテの自動解析は技術的な難易度が高く、実用的な精度を実現するためには多額の研究開発投資が求められる。そこで、本研究提案では、医療現場に直接的なメリットが生じる研究課題に取り組むことによって、現場の協力と今後の追加的な研究開発投資を呼び込み、その過程を通じて実用性の高い電子カルテの自動解析技術を実現する戦略を採る。初年度、我々が今まで模擬カルテを用いて研究開発を進めてきた標準化技術を、国立病院機構の有する広域電子カルテ網(NCDA)上の実カルテへと適用し、技術的な課題を抽出する。2年目には、NCDAを用いて集積した電子カルテに加えて、退院サマリ情報を用いることで、電子カルテの自動要約技術の検討を行う。3年目においては、両技術の統合により、継続的な精度向上の体制を実現するとともに、研究成果を既存の各社電子カルテへと組み込む枠組みを構築する。本研究により、退院サマリの自動要約技術や紹介状の作成支援技術等、医療用の自然言語処理に関連する多彩な応用技術が実現する。これは、医療現場における負担軽減策として極めて効果が期待される。また、こうした応用の発展により、要素技術である電子カルテ上の記載からの自動情報抽出において、継続的な精度向上が実現する。この手法は、電子カルテにおける用語の標

準化技術単独に研究開発投資を行うことと比して、投資効率が極めて高いと考えられる。さらに、こうして医療用自然言語処理技術が発展することにより、大量の電子カルテからの効率的な情報抽出が実現する。これは健康医療政策に資する統計データの収集コストを劇的に低廉化し、今後、政策に求められる様々なエビデンスを継続的に生み出していく基盤となることが期待される。

なお本分担研究では、本研究における「大規模に電子カルテデータを入手できる体制」として、全国の国立病院55施設より年間120万患者の電子カルテ情報を自動収集する診療情報集積基盤(NCDA)を構築し、運用している基盤を用い、本研究目的のためのデータの収集・分析活動を行うためのシステム構築及び運用を行うことを目的とする。

B. 方法

本研究の今年度申請時に記載した研究計画および方法は以下のとおりであった。

【取り組む課題】

本研究では、電子カルテの記載より標準化した情報の抽出を行う技術の研究開発を行う。たとえば、カルテに「精神遅滞」と記載されている場合、この語を解釈することは機械にとって容易である。しかし、カルテの記載は単純ではない。「MR」と記載されていた場合、その語が精神遅滞(Mental Retardation)を指すの

か、僧帽弁閉鎖不全症 (Mitral Regurgitation)か、製薬会社の営業担当 (Medical Representative)かを正確に判別することは機械にとって容易ではない。さらに、「学業成績が芳しくない」と記載されたとき、医学的には文脈より mild MR と解する柔軟性も求められる。カルテ記載は、往々にして略記法や省略が多用される。「腹部：圧痛、反跳痛なし」という記載から、圧痛の有無を判別することは、高度な処理を要する。医療用自然言語処理はこのように難度の高い処理であり、現時点では情報抽出の精度に限界があり、研究開発体制におけるブレークスルーが求められている。

【研究体制】

研究代表者 堀口は、「大規模に電子カルテデータを入手できる体制」として、全国の国立病院 41 施設より年間 80 万患者の電子カルテ情報を自動収集する診療情報集積基盤 (NCDA) を構築し、運用している。この研究基盤を用いることにより、大学病院等のように患者の偏りが生じない多彩な施設からカルテ情報を収集することが出来る。

研究分担者 狩野・森田は、日本語カルテに記載された所見や病名を自動的に標準化するコンテスト (NTCIR MedNLP) を主催する、医療用自然言語処理における我が国の代表的な研究者である。このコンテストは、日本語カルテからの情報抽出精度を競う唯一の学術集会であり、森田は参加者としても首位の成績を収めて来た。狩野は、コンピュータによる大学入

試問題の自動解答に向けた人工知能研究「ロボットは東大に入れるか」プロジェクトにおいて、社会科分野の科目担当リーダーを務めている。両名は、公募研究課題の求める「工学系人工知能研究者・自然言語処理研究者」の役割を果たす。

分担研究者 奥村は、我が国においては数少ない計算機科学分野において学位を取得した医師であり、自然言語処理の医療応用分野で継続的な貢献を果たしてきた。本研究において奥村は、「病名・治療の標準化に詳しい臨床医学系研究者」として、基礎技術研究と応用研究とを橋渡しするコーディネータ役を担う。

分担研究者 岡田は、本研究の対象となる各国立病院カルテを記載する臨床医を取りまとめ、研究開発成果の臨床的有用性の担保に取り組む。

【H28 年度実施状況】

初年度においては、上述の NTCIR MedNLP コンテストを中心に研究開発を進めてきた標準化(正規化)技術を、NCDA 上の実カルテへと適用し、予備的な検証を行うこととしていた。実際には

1, 実データは利用場所が限られている等、研究利用に対して制限が多いことから、実際の診療を行っている医師の協力を得て、実際の診療経過に沿ったダミーデータを作成し、各種研究開発を加速する為の基盤を作成した。

2, 実際の分析を効率的に行うため、セキュリティを考慮した分析環境の整備/構築を行った。

3, 退院サマリと、電子カルテの医師記録

についての関係について、現場医師から情報提供を受け予備的な検討を行った。

【H29 年度計画】

2 年目においては、NCDA を用いて集積した電子カルテに及び、退院サマリデータを対象とした電子カルテの自動要約技術の検証を行う。平成 28 年度に構築した分析環境基盤を活用して、まずは、医師記録や検査データ等入院期間内に生成された情報と、退院サマリデータの差分の検証を行い、医師がどのような思考経路を経て情報処理を行い、文書を作成しているかについての検討を行う。その際、一般的な文章を対象とした自動要約技術の応用に加えて、医療文書を対象とする正規化技術を活用した疾患や病期毎のテンプレートを用いた情報抽出を試みる。この過程は、医療従事者による詳細な手本(教師データ)を用いることで精度向上が望める。その際、平成 28 年度に作成したダミーデータを活用し、本番データの分析環境と平行して各々の施設内での作業を並行して行っていくことで研究活動の効率化を図る。

この研究計画を踏まえ、本年度においては、我々が日本語における医療用自然言語処理の研究コミュニティを形成し研究に取り組んで来た標準化技術を実カルテへと適用することで、カルテからの情報抽出の自動化に向けた検証を行うことを計画した。

その計画の実現に向け、昨年度に引き続き採択後研究者全員で組織する「総括・企画調整班」を作り、月 1 回程度のミーティングを行い、研究の方向性の整理及び各分担任の作成及び役割の決定、進捗の管理及び

調整を行う方法で研究を遂行することとした。また、「総括・企画調整班」以外の分担任についてはそれぞれ責任研究者を決め、その裁量で研究を進める方法をとった。

(尚、総括・企画調整班以外の分担任については総括・企画調整班の活動結果から生まれたものであり、それぞれの班の目的・方法についても C. 結果セクションで記載することとする。)

C. 結果

1. 総括・企画調整班

まず、本研究班はその応募要項の段階からデータ収集に掛かる部分については研究の中に組み込まないことを求められており、データの収集基盤の構築・運営については本研究のカバーする範囲ではない。しかしながら、本研究の前提となる国立病院機構が作成・維持運営する NCDA について本報告書でその概要や意義について記述を行わないとするならば、本報告書の内容の理解に大きな妨げになると考えここに報告を行うこととする。

NCDA の平成 29 年度の概要については参考資料 1 にその概要を資料を添付した。また、実際の病院における SS-MIX データ作成に掛かるシステム仕様についても参考資料 2 に示した。

これらの仕様等のドキュメントについてはその改版履歴も含め、github 上で管理、公開している。

https://github.com/nhoHQ/SSMIX2_support_documents

次に、実際の本分担任研究班の活動について

の報告を行う。本研究班においては、まずは研究者が独立して研究活動を進めるのではなく、本年度10回の研究班会議（うち9回はWeb音声会議）を行い、1つの有機的な研究班として活動が行える環境で運営してきた。

各班会議での調整事項は以下の通りである。

第11回

- 平成29年度計画の確認

第12回

- 各自状況報告
- 研究成果報告書
- 記録の保管について

第13回

- 各自状況報告

研究成果報告書

年間スケジュール

第14回

- CLEF シェアードタスク
- 荒牧班との共同研究班会議について
- 情報共有／相談
- 研究分担の整理
- 「良質な退院時サマリとは？」問題の整理
- ダミーカルテ作成の是非

第15回

- 匿名化について
- 荒牧班との共同研究班会議
- 各自状況報告

第16回

- 匿名化について
- 各自状況報告
- 年度内達成目標の再確認

- 各分担研究状況報告
 - テストデータについて
- #### 第17回

- 各分担研究状況報告
 - 研究基盤の整理
- #### 第18回

- 各分担研究状況報告
- 研究発表の場について
- スケジュール確認

第19回

- 各分担研究状況報告
- 報告書作成について
- 分析環境の整備について報告

第20回

- 各分担研究状況報告
- 報告書作成について

なお、本研究の倫理審査申請書は参考資料3に示した。

なお、NCDA データは国立病院機構が契約するデータセンター内で厳重に管理されている。研究に際しては、このデータベースから研究テーマごとに匿名化したサブセットを切り出し、国立病院機構本部内のオンサイト利用に限っている。以上により、データセットの利用対象と利用目的を厳しく制限することにより、患者個人情報の保護を行っている。それに対応する分析基盤の作成に関して、分担研究班を組織し、昨年度に引き続き堀口・岡田が責任者として活動を行うこととした。（分担研究の結果は後述）

また、昨年度に引き続き、「退院サマリの自動生成に向けたアプローチの検討」というテーマの分担研究を奥村が、「退院サマリの自由記載文の特徴解析」というテーマの分

担研究を森田が、「退院サマリの自動生成に向けたアプローチの検討」というテーマの分担研究を狩野が担当することとした。

2. SS-MIX2 分析用データセットの作成・開発班

研究代表者堀口及び分担研究者岡田は、国立病院機構本部との調整を中心とした基盤構築を行った。昨年度 NDCA データの研究利用に向け、倫理審査申請に加えて、内部規定にて定められている内部委員会の調整を図りともに承認を得た。また、閲覧・解析に特化した自然言語処理用の研究基盤の構築を行った。研究基盤の概念図は図1に示したとおりで、セキュリティを維持しつつ、空間的制約をなるべく少なく研究が進められるようなものになっている。

また、本研究で中心的に使われる医師記録等（経過記録・退院サマリ）については、SS-MIX2 の標準仕様に含まれていないが、JAHIS の提供している仕様を参考に、資料1で示した仕様で NCDA 内に実装することとした。

3. 退院サマリの自動生成に向けたアプローチの検討班

入院患者の退院に際し、医師は入院中に記載したカルテ等の情報から退院サマリを作成する必要がある。この退院サマリを自動的に生成することが出来れば、臨床現場の負担を下げることが出来ると共に、医療の質に貢献することが期待される。そこで、本研究分担では、退院サマリの自

動生成に向けた研究アプローチの検討に取り組んだ。まず、文献調査と医師へのヒアリングに基づき、良質な退院サマリに求められる要件について定性的な検討を行った。同時に、実際の退院サマリを対象とした分析を行い、要約過程に関する知見を整理した。さらに、一般的な文書の要約手法と入院カルテの要約手法について文献調査を行った。

昨年度の研究の結果として、入院カルテの自動要約に向けて、医師が作成する退院サマリの分析枠組みと自動的な退院サマリ生成モデルの双方を兼ね備えた、CASE モデルと証するモデルを構築した。これは、「そもそも退院サマリには何が書かれているのか」という観点より構築されたモデルであり、サマリ中の各文を「カルテに由来するかどうか」という軸と「抽象度が高いか低いか」という軸によって4つのクラスに分類する(図2)。これらのクラスは、退院サマリの分類モデルであると同時に、それぞれ生成に際して固有の処理が求められることから、退院サマリの自動生成に向けた処理モデルとしての性質も有する。以下では、これらクラスの分類軸となる「カルテ記載の有無」、「言及の抽象度・事実度」それぞれを概観した上で、退院サマリの生成モデルとしての展望を記す。

入院治療を受けた患者が退院するのに際して、医師は、記載していた入院カルテを要約して退院サマリを作成する。入院カルテから退院サマリを自動生成することが出来れば、臨床現場の負担を下げることが出来ると共に、さまざまな波及効果が期待される。本研究班は、昨年度よりこの研究テーマに取り組み、昨年度、退院サマリの自

動生成に向けたアプローチを検討した。その結果、退院サマリの自動生成を実現するCASEモデルと称する処理モデルを構築した。

今年度は、このモデルの実証と活用に向けた研究基盤の構築を進めた。まず、研究の推進に際して、自由に研究利用が可能なダミーカルテの整備を行うと共に、アノテーションガイドラインの開発を試みた。また、CASEモデルの実証に求められる「カルテ中センテンスの抽象度分類器」の準備を進めた。さらに、退院サマリの生成システムのユーザーインターフェースを設計し、プロトタイピングを行った。

入院カルテの自動要約に向けた研究は、医療用自然言語処理研究において未開拓な領域であり、まずは研究基盤の構築と合わせた研究方法論の確立が求められる。昨年度、今年度の研究を通じて、研究アプローチの検討と研究基盤の整備が進んだ。

4. 退院サマリの自由記載文の特徴解析班

退院サマリの各文について、元になった入院カルテと比較をすることで、その文が入院カルテからそのまま抜き出された文なのか、文や文節などを組み合わせて書かれた文なのか、それとも入院カルテの記載を解釈して新たに生成された文なのか、を決定する。

退院サマリの自動生成技術の実現を目指し、そのために必要な2種類の調査研究を行った。退院サマリを生成するためには、どのような退院サマリを生成しなくてはならないかを示す「理想的な退院サマリ」の定義が求められる。そこで、理想的な退

院サマリについて言及をした国内外の研究の文献を調査した。また、自動生成をした退院サマリが意味のある退院サマリとして成立しているかは評価が必要である。そこで、要約文や自由記載文の自動評価手法に関する国内外の研究の文献を調査した。PubMedや医中誌などを用いて文献を検索し、理想的な退院サマリに関する文献を51報（英文48報、邦文3報）、要約文や自由記載文の自動評価手法に関する文献を243報（英文：148報、邦文95報）得た。理想的な退院サマリに関する文献では、記載されることが望ましい項目は何か、各項目が実際の退院サマリに記載されているか、項目ではなく文章としてどのような特徴を有していることが望ましいのか、などが言及されていた。一方、要約文や自由記載文の自動評価手法は、評価の際に正解となる退院サマリが用意できる場合に利用できる手法と、用意できない場合に利用できる手法とに分けることができた。研究で使用する場合には前者が適用できるが、実臨床で使用する場合には後者が必要になると考えられる。次年度は、本研究で得られた知見を整理し、他分担グループと連携をして退院サマリの自動生成および評価に取り組む。

5. 退院サマリの自動生成に向けたアプローチの検討班

電子カルテの実データに対し処理を行うには、まず匿名化が必要となる。具体的には、個人氏名、年齢、住所、日付、医院名など個人を特定しうる情報の自動抽出である。

匿名化の研究は長らく行われているもの

の、特に日本語医療分野の匿名化は利用可能なリソースが少ないこともあり、研究の数は限られている。利用可能なリソースとしては、NTCIR MedNLP 匿名化タスクで配布された模擬カルテコーパスとそこに付与されたアノテーションが挙げられる。このとき評価に使われた正解アノテーションは期間限定で利用不能となっており、残念ながら直接的な比較を行うことはできない。

そこで、研究班内で作成されたダミーカルテを対象に、新たに匿名化のための正解アノテーションを付与し、学習及び評価に用いた。付与するアノテーションは基本的に MedNLP タスクにおけるものと同種とした。すなわち、年齢・個人名・医院名・性別・時間表現である。

これらのデータを用いて、自動匿名化ツールのプロトタイプ実装を行い、性能を検証した。手法としては、ルールベースのものと機械学習によるもの、およびそれらの混合を試みた。前述の理由から、MedNLP タスクにおける先行研究との直接比較はできないが、概ね同等程度の性能が得られたと考えられる。ただし、いずれのデータセットも模擬カルテの作成コストが大きく、サンプル数が不足している。そのため、end-to-end の機械学習のみでは性能が不十分であり、ルールベースの併用が必要である。今後はこれをより大規模なデータに適用し、実用化を図る。

また、サマ리를要約ととらえると、一般的な自動要約技術が適用できそうにも思われる。多くの自動要約は、トピックの統計的な解析を行ったうえで、文書中で重要なトピックが含まれるものを残す、という手

法が骨格になっている。しかし退院サマリでは、統計的に重要でない、文書集合中で共通して頻出するトピックであっても、サマリとして残すべきことが多々ある。

また、入力にあたる電子カルテの文章中にない文章や表現が、サマリにどのくらい含まれているかという問題がある。入力のサブセットでよいのであれば、切り貼りの範囲内におさまるが、現実には言い換えに始まり内容的にも新規な文章の挿入がありうる。

分担研究のデータによると、入院までの経過については7割以上のサマリ文がカルテの文章ほぼそのままであった。このことは、医師がサマリを作成する際に文の複製を使用しており、分量的な減少もあまりみられないことから、内容的にもあまり変更を必要としていないことを示唆している。ただし、入院前の記述は他の医師からのお願いの形式をとっており、そのままでは主語や言葉遣い、時制などが不適切なので、そうした部分の変換が必要かもしれない。

入院中の経過については、2～3割程度がそのままの文であった。入院中のカルテの記載は文を完成させず断片的なスタイルのことが多く、一方サマリではきちんとした文にするため多かれ少なかれ文生成の要素が必要と思われる。また、医師本人の記録なので、振り返ることで要素を追加したり、整理したりすることが想像される。

このようにさまざまな要素があるが、現実には診療科や疾患のタイプによって、類似性の高いものとそうでないものがありうる。そこで、国立病院機構内の実際のカルテを対象にクラスタリングを行い、カルテに記載された ICD や DPC といったラベルを含め

て類似性の解析を行った。結果、ある程度の類似クラスタの形成が見られたが、より安定した分析にはもっと多くのデータが必要と考えられるため、今後はさらにデータ量を増やし分析を進める。

D・E.考察及び結論

医療用情報システムの研究開発においては、医療現場に直接の恩恵が及ばないゴールが設定されることで、研究開発が現場のニーズから乖離するとともに、継続した開発投資に繋がらない悪循環が往々にして生じてきた。本研究提案は、医療現場における負担軽減策として期待が大きい退院サマリの自動要約技術の開発を目指す。これにより、紹介状の自動作成技術等、電子カルテの自動解析技術に関連する継続的な研究開発投資の実現が期待される。この研究開発サイクルを確立することにより、要素技術である電子カルテ上の記載から自動情報抽出における継続的な精度向上が期待される。

こうして確立する医療用自然言語処理技術は、大量の電子カルテからの効率的な情報抽出を実現し、健康医療政策に資する統計データの収集コストを劇的に低廉化することが期待される。とりわけ、様々な傷病や治療に関して、既存のDPCやレセプトには表れてこない深遠な実態を明らかとし、医療の質向上・均てん化・各種医療技術の臨床開発に必要なエビデンスを生み出していくことが期待される。

また、国立病院機構の有する広域電子カルテ網は、各病院が独自に調達した電子カルテベンダー主要6社を網羅している。本

研究によって、大口顧客としての交渉力を背景としたこれら主要ベンダーへの研究開発成果の技術移転が期待される。これらは、医療の情報化を進める厚生労働行政にとって、新たな政策手段の実現をもたらす

加えて、本研究で作成した入院カルテ・退院サマリの高品質な個人情報を含まない模擬のデータセットは、研究成果として公表することにしており、電子カルテの現物を持たない情報系研究者が、容易に本領域の発展に貢献できる環境を提供出来る点も大きな成果になるものと考えている。

今年度の研究においては、NDCAデータの利用に際した倫理審査の関係で、実カルテを対象とした解析は3月以降に限られる制約が生じた。しかしながら、高品質なテストデータの確保が可能であったことから、研究のコアである医療用自然言語処理部分の検討はほぼ計画通りに進めることが出来た。電子カルテの自由記載部分を自動解析する多施設構成での大規模データを対象とした研究としては、本邦初の試みとなる。今後も本計画に沿った研究を遂行していきたい。

F. 研究発表

1. 論文発表

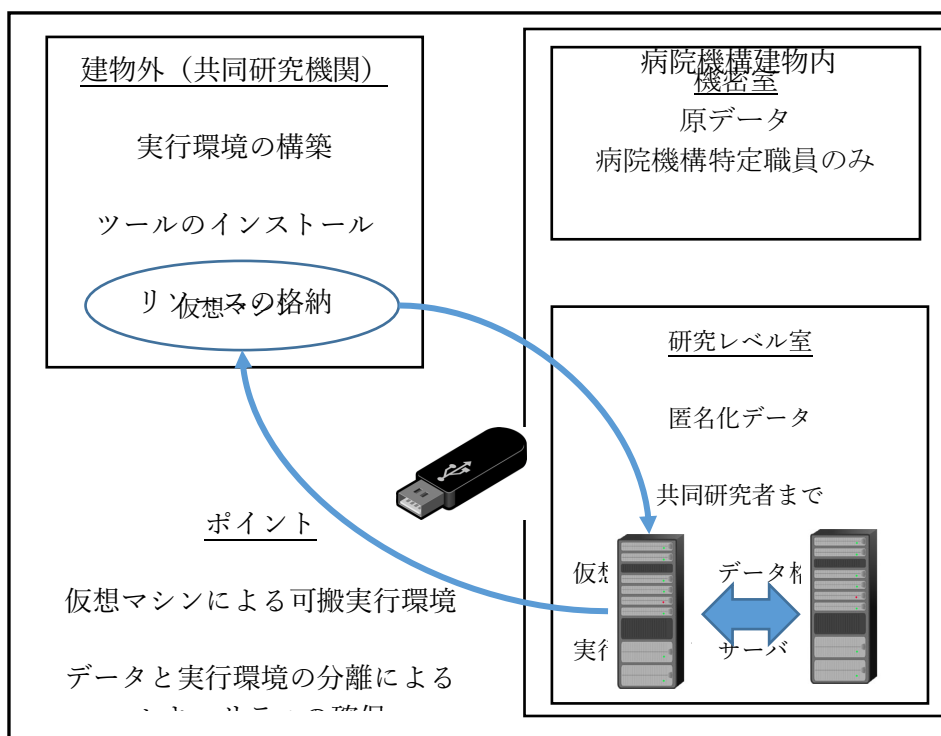
堀口裕正: 国立病院機構のデータベースを用いた臨床研究 *Progress in Medicine* 38(2):17-20 2018

堀口裕正: NCDCA の現況と成果および今後の展望 月刊新医療 2018年2月号

2. 学会発表

なし

図 1



資料 1 NCDA における医師記録等の仕様書

趣旨

本事業では、各社の **SS-MIX2** モジュールの拡張ストレージへの出力機能を利用し、以下の情報を出力することを求めている。その際、**SS-MIX2** 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d (以下、ガイドライン) に記載している仕様に対応していること。また、トランザクションストレージ、インデックスデータベースも同時に生成すること。

- 経過記録
- 退院時サマリー
- 診療情報提供書

以下に仕様を示す。


```
|-- 拡張ストレージ ルートフォルダ
  |-- 患者 ID 先頭 3 文字
    |-- 患者 ID 4～6 文字
      |-- 患者 ID
        |-- 診療日
          |-- データ種別
            |-- コンテンツフォルダ
              |-- 主文書ファイル
```

診療日

特に指定しない。

データ種別

ガイドライン P4 (4) 「データ種別フォルダ」について に則ること。

```
[ローカル文書コード]^ローカル文書名称^[ローカルコード体系コード]^標準文書コード^
標準文書名称^標準コード体系コード
```

以下のように標準コードに対しローカルコードが複数あることは許容される。

```
L12345^入院診療録^99ZZZ^11506-3^経過記録^LN
```

```
L12346^外来診療録^99ZZZ^11506-3^経過記録^LN
```

コンテンツフォルダ

ガイドライン Ver.1.2d P5 (5) 「コンテンツフォルダ」について に則ること。

```
患者 ID_診療日_データ種別コード_特定キー_発生日時_診療科コード_コンディションフ
ラグ
```

いずれの文書も削除は想定していないが、電子カルテシステムによっては修正はあり得ると考える。その場合、ガイドライン P6 ④修正が発生する場合 に則り改版すること。

主文書ファイル

XML CDA R2 で出力すること。XML ファイル以外に画像ファイルや CSS ファイル等を出
力してもかまわない。

HEADER 部

いずれの文書も JAHIS 診療文書構造化記述規約 共通編 Ver.1.0 に則ること。

P27 6.3.11.検査・診療等行為 "documentationOf/ServiceEvent" によると、documentationOf
の制約・多重度は 0..1 となっているが、経過記録、退院時サマリについてはこれを 1..1 と
読み替えること。

経過記録は serviceEvent classCode(サービスイベントクラスコード)を ENC(診察)とし、
effectiveTime(実施日)は low value、high value とともに記録タイミングを出力すること。

退院時サマリは serviceEvent classCode(サービスイベントクラスコード)を ACCM(入院、
滞在)とし、effectiveTime(実施日)は low value に入院タイミング、high value に退院タイミ
ングを出力すること。

タイミングの粒度は日以上であれば良い。

BODY 部

診療情報提供書は、日本 HL7 協会 患者診療情報提供書 規格 Ver.1.00 に則ること。

診療情報提供書以外は、XML の文法に則ること。

参考資料

1. NCDA データベースの説明資料

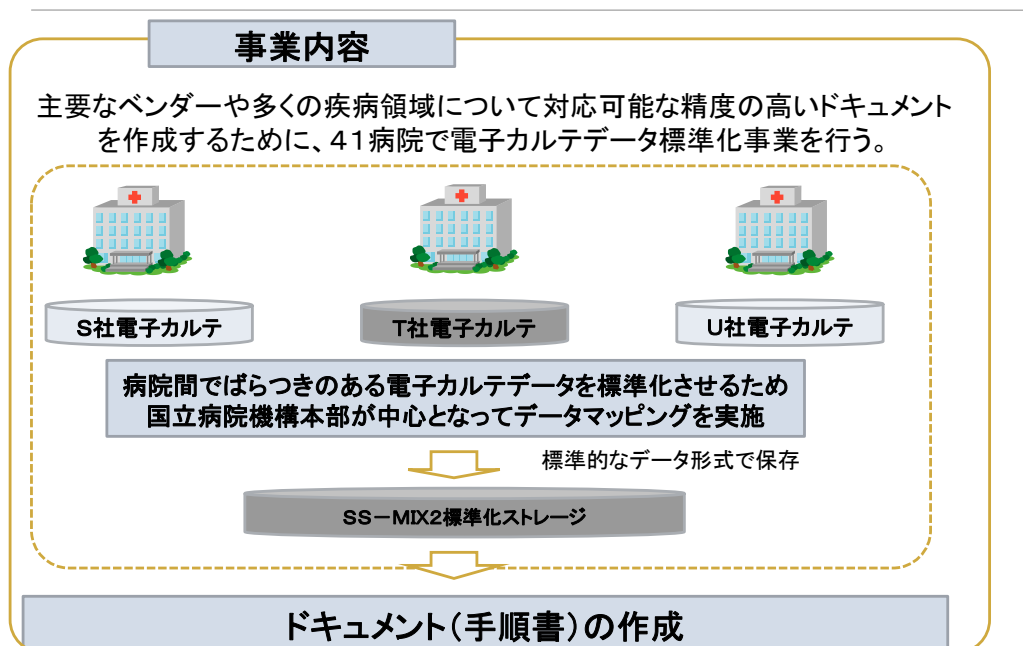
国立病院機構診療情報集積基盤について (NCDA:NHO Clinical Date Archives) ～電子カルテデータの標準化～

国立病院機構本部
情報システム統括部

NCDA(補助金事業)の事業背景

- 平成26年6月24日に閣議決定された「世界最先端IT国家創造宣言」では、地域を越えた国民への医療サービスの提供等を可能とする医療情報利活用基盤の構築を目指し、医療情報連携ネットワークについては、電子カルテを含めたデータやシステム仕様の標準化等を行い、平成30年度までに全国への普及・展開を図ることとされている。
- しかしながら、電子カルテについては、ベンダー毎で開発が行われ、各病院が使いやすいようにカスタマイズされるなど、電子カルテデータの形式が標準化されないまま普及したことから、電子カルテ上で使用されている病名や医薬品等のコードがベンダーや病院で異なり、標準化の課題となっている。
- 国立病院機構の『電子カルテデータ標準化等のためのIT基盤構築事業(H26補正:13.0億円)』では、このような問題を解消するため、各病院の電子カルテデータを厚生労働省の定める標準コードに紐付けするデータマッピングを行い、SS-MIX2規格(標準化ストレージ機能)を用いて電子カルテデータの標準化を実施し、その工程を示したドキュメント(手順書)を作成・公開することを目的としている。

補助金事業の概要（課題・目的等）



2

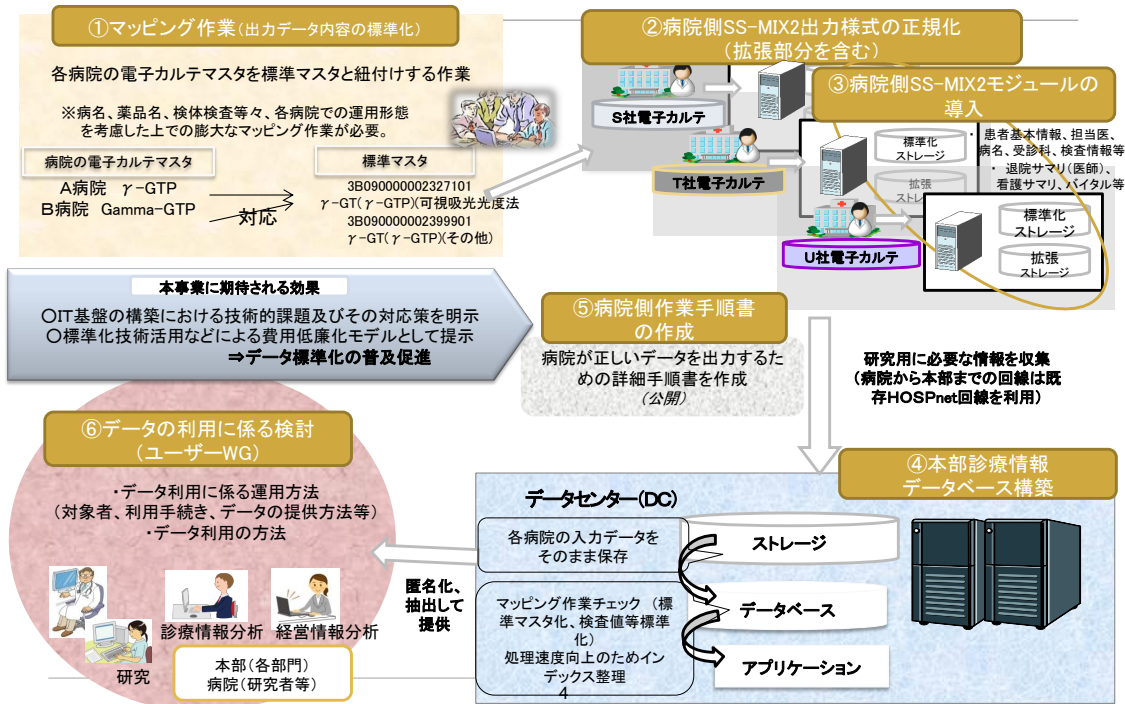
国立病院機構のDB事業概要（プロジェクト概要）

方針

主な作業区分	内容
①マッピング作業（出力データ内容の標準化）	対象41病院を選定し、データマッピング作業を実施する
②病院側SS-MIX2出力様式の正規化（拡張部分を含む）	全てのSS-MIX機能（メッセージ）に対応できるように、モジュールを各ベンダで正規化（入力値の正規化・フルセット化等）する。併せて標準仕様以外の拡張データ（バイタル等）が出力できるようにする
③病院側SS-MIX2モジュールの導入	①で選定した対象病院に②で作成したSS-MIX2モジュールを導入する
④本部診療情報データベースシステム構築	データを収集する仕組みを検討し、外部データセンターにデータベースを構築する
⑤作業手順書の作成	本プロジェクト終了後、各病院がSS-MIX2を効率的に導入できるように、SS-MIX2モジュールを導入するベンダが作業手順書を作成する（手順書は公開予定）
⑥データ利用に係る検討（ユーザーWG）	システム機能とユーザーの要望について調整するデータベースの利用に係る規定（プロセスやルール）や具体的なデータ利用方法を検討する

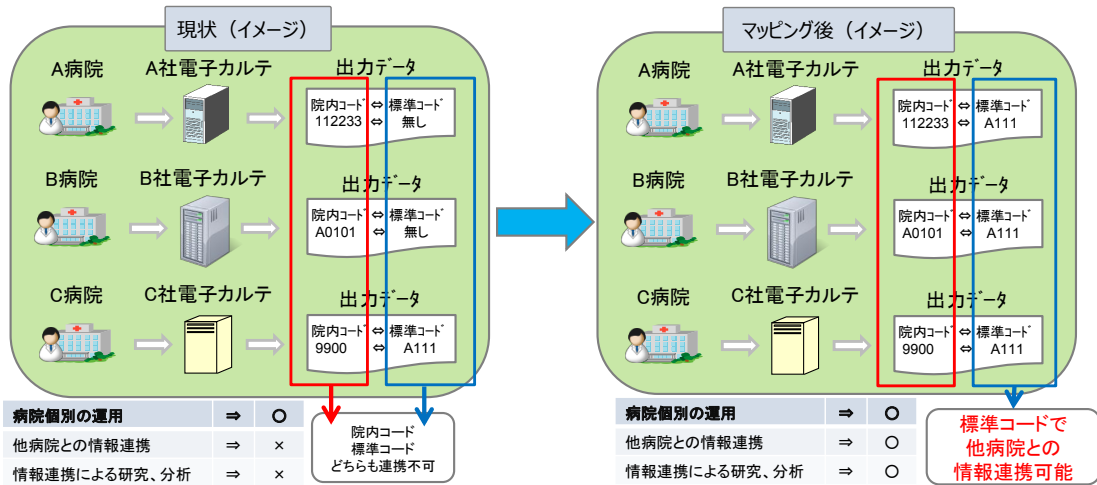
3

SS-MIX2を用いた診療情報データベース構築プロジェクト 作業区分①～⑥

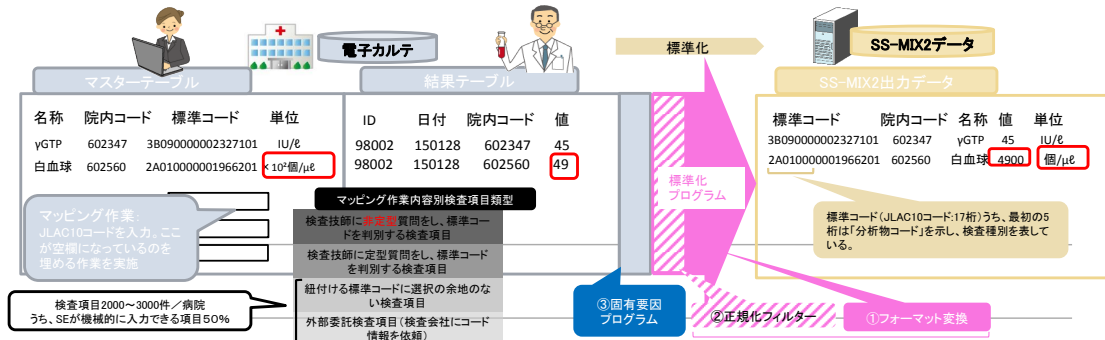


病院におけるマッピング作業

- 院内コードと標準コードを紐付ける対応表を作成します(マッピング作業)。
- 病院毎に異なる院内コードを、標準コードに変換することにより、他病院と連携した診療情報の分析等が可能になります。



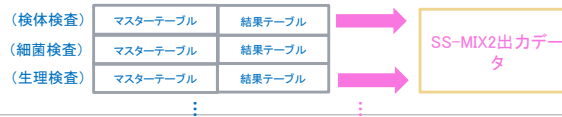
データ標準化のイメージ(SS-MIX2出力)



- ①フォーマット変換: 電子カルテのフォーマットをSS-MIX2フォーマットに変換する
- ②正規化フィルター: 例えば、同一検査で単位が混在している場合、標準とされる単位に変換する(白血球検査で「個/μℓ」と「×10⁶個/μℓ」が混在している場合、標準の単位が「個/μℓ」であれば、院内電子カルテ上で「×10⁶個/μℓ」の単位にて表されている「値」については、×100してSS-MIX2出力データとする)。他には、使用している文字コードが違う場合、標準とされる方にあわせて変換する、など。
- ③固有要因プログラム: 病院独自の検査表示など実施している場合、出力時に標準に合うように変換する。

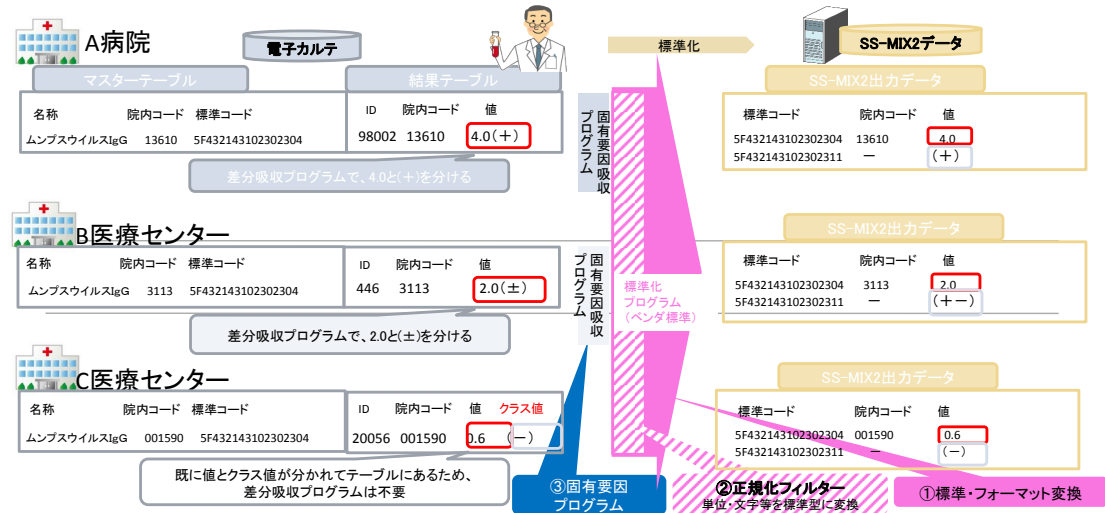
①②は標準化プログラムとして電子カルテベンダーにて開発。
③は各病院における電子カルテ導入業者が開発。

電子カルテデータからSS-MIX2データへ変換する際、対象となるカテゴリは(検体検査)など数十個存在。
すべてのカテゴリを変換する標準化プログラムについて、確認・検証する必要がある。



6


SS-MIX2変換プログラムの構成



- ①②の標準化プログラムは他の施設でも使用する汎用的なもの。
 - ③の固有要因プログラムについては病院固有のもの。
- ベンダー側が構築する標準化プログラム①②に固有要因プログラム③の機能が含まれていると、その病院でしか使用できない(汎用化されていない)ことになり、普及促進を図る手順書としての品質は不可。
→複数病院で標準化プログラムを運用して、それが汎用的なものであること(病院固有の変換機能が入っていないこと)を確認する必要がある。
※①②③のプログラムの著作権はベンダーにあるため、コード等中身を見ることができない。よってNHOが結果により確認する必要がある。

7

事業の成果(標準化の普及促進関係) H28.3時点

- 最新のSS-MIX2Ver1.2cに完全準拠しているモジュールを41病院に導入
 - SS-MIX2 Ver1.2cモジュールの導入
 - SS-MIX2に完全準拠しているモジュール
 - HOTコード・JLAC10・ICD10など標準コードを全面的に導入・活用
- 
- 従前のモジュールで課題となっていたベンダー毎の表記ゆれ等の問題が解決され、データ形式の標準化が可能となる。
 - 本モジュールは主要カルテベンダー7社から他の医療機関にも(有償にて)提供可能。
 - 他の医療機関が厚生労働省標準規格に準拠(SS-MIX2・標準コード等)したシステムを導入するに当たり、当該事業で作成したドキュメント(手順書)を活用することにより、専門的な知識を要することなく、簡便に導入することが可能となる。

8

病院側の負担について

- 本事業の病院募集時において試算したところ、病院へのモジュールの導入費用900万~1000万、翌年度以降の(保守・利用料等の)病院負担額は年額100万円以内と設定
- 初期投資についてはベンダー側初期提案額から1~2割の低減を行った。(平均700万円台)
- 本事業内で低廉化について調整したところ全病院年額20万円台で運用できることで調整がついた。
- 電子カルテ更新時のコストについては今後とも交渉を行っていくが、さらなる低廉化がはかれるよう努力していく。

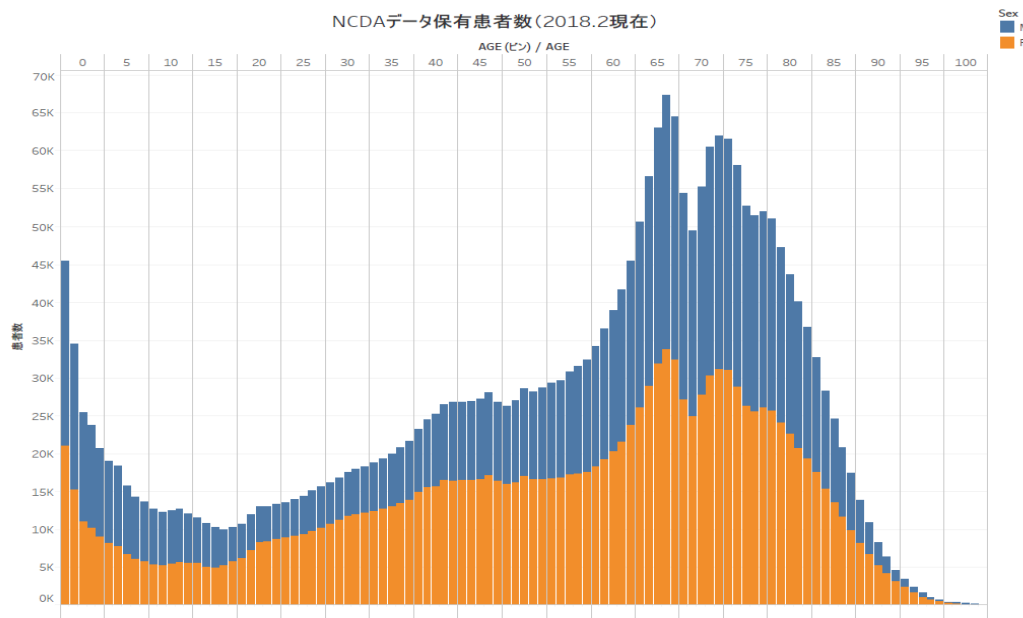
9

NCDAの現状

- ・平成26年度補助金「電子カルテ標準化等のためのIT基盤構築事業」にて6ベンダー41病院で事業開始(平成28年1月～)
- ・平成29年度末に7ベンダー58病院に拡大
- ・年間患者数約120万人規模のデータ基盤
平成30年2月末時点で約160万人のデータを蓄積
(患者数は実患者数、外来データ含む)
- ・一般臨床レベルのデータクオリティはある。
(例えば、検査データは、全て検査値が統一されアーカイブされている。)
- ・現在安定して稼働中

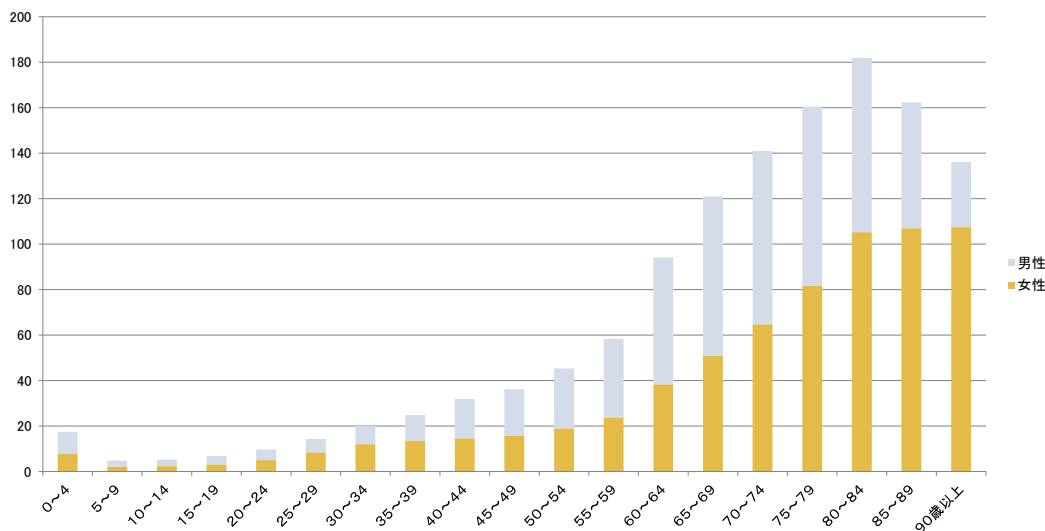
10

NCDAに集積している患者数(年齢・性別)の状況



12

(参考)年齢・性別推計入院患者数(病院)



(単位:千人)

(出典)H26厚生労働省患者調査

13

NCDAと災害診療記録プロジェクト概要①

1 事業名称と予算

◆事業名称:

平成28年度地域診療情報連携推進費補助金

「電子カルテによる『災害診療記録』電子フォーマット自動出力実証事業」

◆予算総額: 2.5億円 (繰越) ※本年度中の執行期限

- ・データセンターにおけるデータ提供モジュール構築に必要な経費
- ・SS-MIX2モジュールのバージョンアップ等に必要な経費 ※病院の電子カルテに関係する部分
- ・手順書作成に必要な経費

◆事業実施主体: 独立行政法人国立病院機構 (単独)

2 事業の背景

○大規模災害時において、災害対策本部(都道府県)が被災地の医療概況を把握し、適確な医療支援活動を展開するうえで、極めて重要な情報は「**疾病別症例数**」等の集計情報であるが、それを迅速に集計する手法の確立が課題。

○この課題に対し、東日本大震災を契機にして「災害時の診療録のあり方に関する合同委員会(※)」が設置され、**災害時の標準的記録フォーム**といえる「**災害診療記録**」が作成されています(熊本地震で初めて運用開始)。

(※)日本医師会、日本集団災害医学会、日本病院会、日本診療情報管理学会、JICA

NCDAと災害診療記録プロジェクト概要②

3 事業の内容

- 電子カルテデータ標準化等のためのIT 基盤構築事業（NCDA）の機能をバージョンアップし、様々なベンダの電子カルテから、自動的に「災害診療記録」電子フォーマット出力が可能となるように開発及び検証を行い、更に導入手順書を公開することを通じて、災害発生時の適確な医療支援活動の展開に役立つもの。

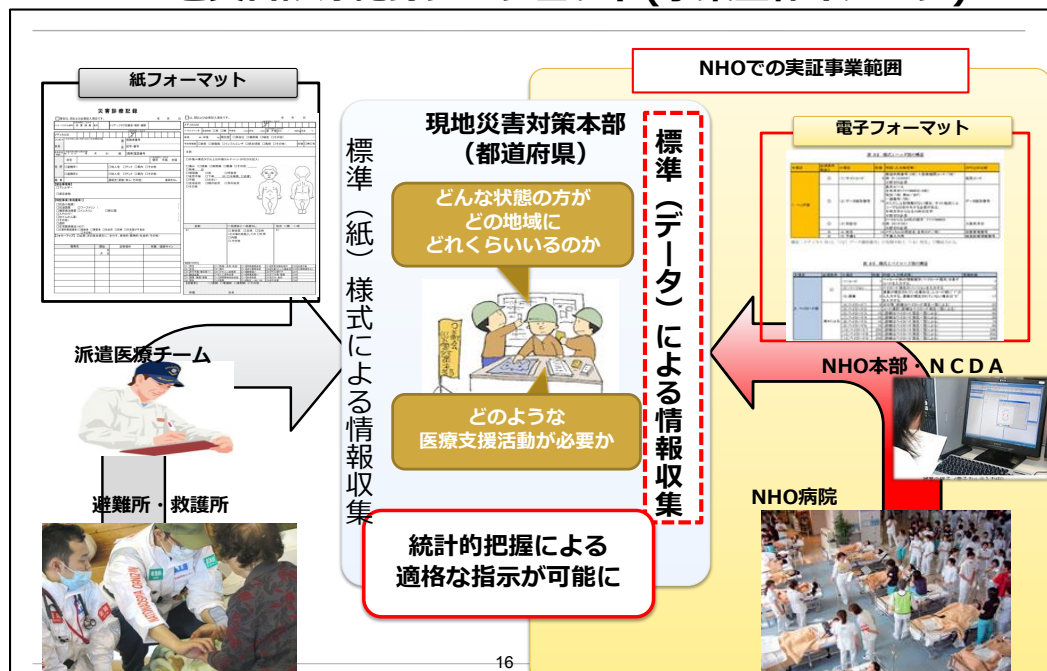
4 国立病院機構で実施する理由

国立病院機構は、

- ✓ 災害対策基本法に基づく指定公共機関
- ✓ NCDAを活用し、効率的に災害診療記録を収集する仕組みの構築が可能
- ✓ 災害診療記録電子フォーマットの普及啓発が可能
(事業の成果のドキュメント化(手順書の作成等))

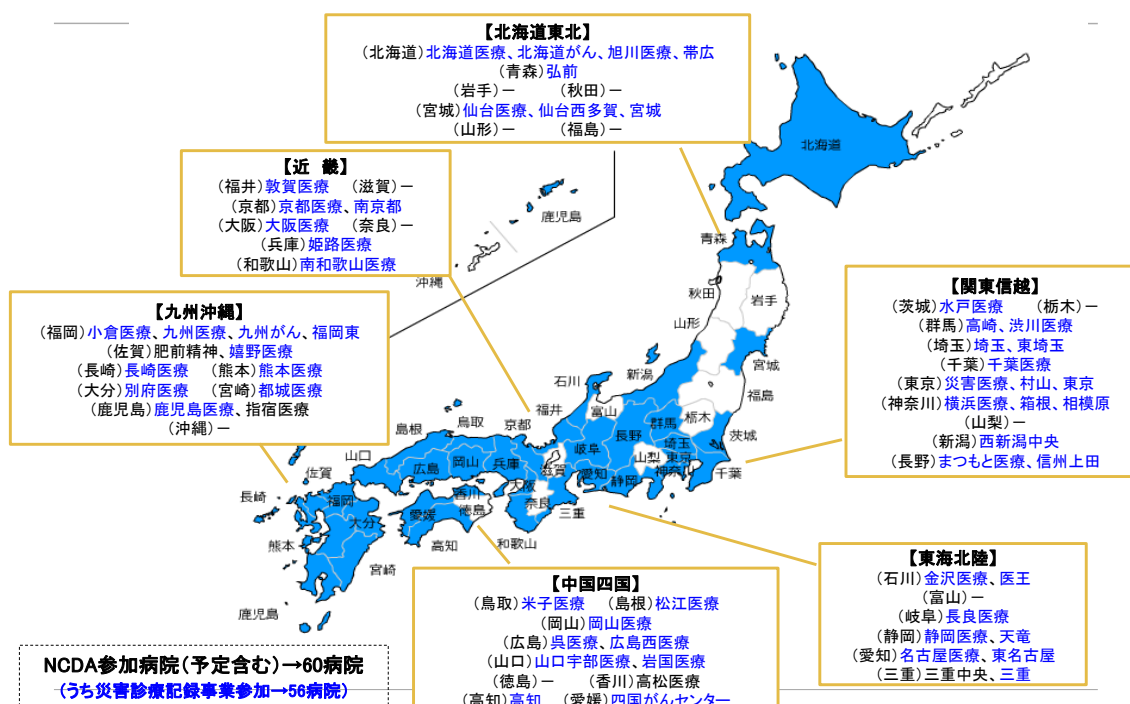
15

NCDAと災害診療記録プロジェクト(事業全体イメージ)



16

(参考)NCDA参加病院の状況



災害診療記録プロジェクトの成果

✓標準的な情報収集における情報の種類の拡大

ONHOのNCDAにおいて現在収集している標準ストレージ内項目36種類(全て)とバイタル情報に加え、経過記録・退院時サマリについて収集できるようにするための病院向けモジュール改修を実施

○JAHIS標準に準拠した形で定義(HL7 V3. 0 CDAベース)

○本研究費では開発・検証までを実施

○NCDA参加7ベンダーのモジュールの開発及び検証環境下での検証は終了

NCDAにおける個人情報取り扱い

1. 患者同意

- 病院に掲示されている「個人情報の利用目的」に「国立病院機構診療情報分析基盤での利用」を追加。
- 併せて、ポスター・ちらしでの周知を開始
- 患者の利用不可の申出には対応できるシステムとなっている

2. 法令対応

- 個人情報保護法・独立行政法人における個人情報保護法、ガイドライン等に適切に対応
- 研究の倫理指針に適切に対応
- 次世代医療基盤法にも適切に対応していく

20

NCDAの利活用について

- 患者に明示した個人情報の利用目的の範囲内で利活用を進める
- 利活用に際しては「利活用要項」を定め、それに従って利用を行う
- 利活用要項の骨子は以下の通り
 - NHO本部内に、データベース利用審査委員会を設置し、データ利用について審議。
 - 利活用は匿名化後が原則
 - 研究における利用
 - 本要綱を遵守するとともに、倫理規定等の研究に関連する法令やルールを遵守する

21

2. NCDA システム仕様書

SS-MIX2 を用いた診療情報データベース構築の為の SS-MIX2 モジュール技術仕様書

1. システム要件

国立病院機構の各病院にて「国立病院機構診療情報分析基盤(NCDA)」に参加する為に調達する SS-MIX2 モジュールの機能は以下の通りである。但し、本体の電子カルテシステム等の仕様上、作成が不可能であるものについては作成を要しない。その場合、何が不可能かを導入標準作業手順書に記載すること。

1.1 SS-MIX2 Ver.1.2d 機能

SS-MIX2 Ver.1.2d に準拠することとして、以下の機能を有すること。

- 日本医療情報学会発行の「SS-MIX2 標準化ストレージ構成の説明と構築ガイドライン Ver.1.2d」、「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d」、「SS-MIX2 標準化ストレージ仕様書 Ver.1.2d」、「標準化ストレージ仕様書別紙：コード表 Ver.1.2d」、「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d 別紙：標準文書コード表」に記載している仕様に対応していること。(尚、当初 Ver.1.2c 準拠としていたが、標準ストレージ部分では Ver.1.2c からの変更点について影響がないため Ver.1.2d 準拠ということとした。)
- 標準化ストレージ、拡張ストレージ、トランザクションストレージ、インデックスデータベースの4つのファイルを生成すること。
- 標準化ストレージにはデータ種別として 36 種のデータを出力すること。

(表 1-1 標準化ストレージ格納データ)

No	データ種別	種別名称	HL7 メッセージ型
1	ADT-00	患者基本情報の更新	ADT^A08
2	ADT-00	患者基本情報の削除	ADT^A23

No	データ種別	種別名称	HL7 メッセージ型
3	ADT-01	担当医の変更	ADT^A54
4	ADT-01	担当医の取消	ADT^A55
5	ADT-12	外来診察の受付	ADT^A04
6	ADT-21	入院予定	ADT^A14
7	ADT-21	入院予定の取消	ADT^A27
8	ADT-22	入院実施	ADT^A01
9	ADT-22	入院実施の取消	ADT^A11
10	ADT-31	外出泊実施	ADT^A21
11	ADT-31	外出泊実施の取消	ADT^A52
12	ADT-32	外出泊帰院実施	ADT^A22
13	ADT-32	外出泊帰院実施の取消	ADT^A53
14	ADT-41	転科・転棟(転室・転床)予定	ADT^A15
15	ADT-41	転科・転棟(転室・転床)予定の取消	ADT^A26
16	ADT-42	転科・転棟(転室・転床)実施	ADT^A02
17	ADT-42	転科・転棟(転室・転床)実施の取消	ADT^A12

No	データ種別	種別名称	HL7 メッセージ型
18	ADT-51	退院予定	ADT^A16
19	ADT-51	退院予定の取消	ADT^A25
20	ADT-52	退院実施	ADT^A03
21	ADT-52	退院実施の取消	ADT^A13
22	ADT-61	アレルギー情報の登録／更新	ADT^A60
23	PPR-01	病名（歴）情報の登録／更新	PPR^ZD1
24	OMD	食事オーダー	OMD^O03
25	OMP-01	処方オーダー	RDE^O11
26	OMP-11	処方実施通知	RAS^O17
27	OMP-02	注射オーダー	RDE^O11
28	OMP-12	注射実施通知	RAS^O17
29	OML-01	検体検査オーダー	OML^O33
30	OML-11	検体検査結果通知	OUL^R22
31	OMG-01	放射線検査オーダー	OMG^O19
32	OMG-11	放射線検査の実施通知	OMI^Z23

No	データ種別	種別名称	HL7 メッセージ型
33	OMG-02	内視鏡検査オーダー	OMG^O19
34	OMG-12	内視鏡検査の実施通知	OMI^Z23
35	OMG-03	生理検査オーダー	OMG^O19
36	OMG-13	生理検査結果通知	ORU^R01

「SS-MIX2 標準化ストレージ構成の説明と構築ガイドライン Ver.1.2d p11」

1.2 拡張ストレージへの出力機能

現在の SS-MIX2 モジュールでオプションとして既に導入している拡張ストレージへの出力機能は、そのまま提供すること。また、1.3.0 で規定する出力を行うこと。

1.3 NHO 対応としての設定

1.3.0 拡張ストレージへの出力機能

各社の SS-MIX2 モジュールの拡張ストレージへの出力機能を利用し、以下の情報を出力すること。その際、日本医療情報学会発行の「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d」に記載している仕様に対応していること。また、トランザクションストレージ、インデックスデータベースも同時に生成すること。

No	データ種別	種別名称	HL7 メッセージ型
1	L-OBSERVATIONS^OBSERVATION S^99ZL01	バイタル検査結果	HL7 V2.5 ORU^R30
2	^ (ローカル名称) ^^11506-3^経過記	診療録 (外来/入院含)	HL7 CDA R2

No	データ種別	種別名称	HL7 メッセージ型
	録^LN	む)	
2.1	^ (ローカル名称) ^^34108-1^外来診療録^LN	診療録 (外来) (入院・外来が別の場合)	HL7 CDA R2
2.2	^ (ローカル名称) ^^34112-3^入院診療録^LN	診療録 (入院) (入院・外来が別の場合)	HL7 CDA R2
3	^ (ローカル名称) ^^18842-5^退院時サマリー^LN	退院時サマリー	HL7 CDA R2
4	^ (ローカル名称) ^^57133-1^紹介状^LN	診療情報提供書	HL7 CDA R2

1.3.1 バイタル検査結果通知の出力

(1) バイタル検査結果通知のデータを、別紙の形式で拡張ストレージに出力する。尚、「診療日」に出力する日付は **OBX-14** トランザクション日時 (測定した日) とする。

(2) ファイル作成の単位は、データの格納構造として日付の下にあるため、最大でも一日分が1ファイルにまとまっている形とする。一日の中で測定のたびに作成するのも良い。一日1ファイルなら、特定キーは測定日を出力する。一日に複数回のデータを出力する場合は、特定キーに測定日の時間まで (YYYYMMDDHH) 出力すること。

1.3.2 バイタルデータの項目及び形式等

(1) バイタルデータとして取得する項目は、「拡張期血圧、収縮期血圧、脈拍数、呼吸数、体温」の5項目とする。

(2) **OBX-3** 検査項目に出力するコードは **JLAC10** コードとする。バイタルデータを参考に適切な **JLAC10** を選択すること。

(3) 上記以外の項目を **SS-MIX2** に出力することは問題ないが、今回の対応では扱わない。但し、今後の検討で仕様として扱うことになる場合は、**JLAC10** コードを基準とした標準コードを必須とすることを想定している。この今後想定される検査項目は別表として提供する。

1.3.3 標準コード変換機能

SS-MIX2 データの出力に際しては、コードのマッピング表などに従って、院内のローカルコードを厚労省が定める標準コードに変換する機能を有すること。またマッピング表については、容易にその内容を変更できるマスターメンテナンスプログラム等の機能を有すること。

JLAC10 コード、**JANIS** コード、**HOT** コードについては、機構病院が **NCDA** 事業に参加する場合においては機構から提供する。

1.3.4 標準化ストレージにおける文字コードについて

メッセージの文字コードについては、「標準化ストレージガイドライン」で示されているとおり、1バイト系文字は **ISO IR-6 (ASCII)**、2バイト系文字は **ISO IR87 (JIS X 0208 第一水準、第二水準)** とする。ただし現実には上記以外の文字コードが電子カルテシステムに登録されている可能性があるため、以下のように対応することとする。

- 1 半角カナ文字 → 全角カナ文字に置き換えて **SS-MIX2** に出力する。
- 2 外字 → ■で置き換えて **SS-MIX2** に出力する。
- 3 環境依存文字については変換表を機構より提供するのでそれにより変換して **SS-MIX2** に出力する。

1.3.5 単位の文字表記の統一

SS-MIX2 データの出力に際して、臨床検査データの **OBX** セグメントの **6** フィールド目の単位の文字表記を統一すること。

【単位の文字表記の統一ルール例】 **ASCII** コードで表記すること

- ・かける → . (ドット)
- ・乗 → * (アスタリスク)
- ・ μ → u (小文字ユー)
- ・語尾に名称 → () で

- °C → cel
- ‰ → permil
- 個 → pcs

【上記ルールの適用例】

- mL → mL (ASCII コード)
- $X10^2/\mu\text{l}$ → $.10^2/\text{uL}$ (かける、乗、 μ)
- /HPF → /(hpf) (語尾に名称)

1.3.6 単位変換機能

SS-MIX2 データの出力に際して臨床検査データの単位に関しては、JLAC10 コードごとに、機構が定める単位に変換を行った上で SS-MIX2 データを生成すること。尚、JLAC10 コード別の単位表は別途機構から提供する。単位表は「SS-MIX2 標準化ストレージ仕様書 Ver.1.2」にも別表として添付する。

【単位変換例】

JLAC10 コード	数値	単位	→	JLAC10 コード	数値	単位
1A02500000127201	10.5	mg/l	→	1A02500000127201	1.05	mg/dL

1.3.7 計測値等の表記方法について

(1) 定性値・検出限界以下・検出限界以上の表記

- OBX (検体検査結果) セグメントの5フィールド目 (検査値) に検査結果を記述する場合、現在そのデータ形式は OBX-2 フィールドの説明にあるように NM 型、ST 型、CWE 型のうちいずれかの形式で記述することとなっている。
- 今回の仕様では、定性値・検出限界以下・検出限界以上のデータについては、SN 型の表現方法を用いて SN 型の“A”を” “ (スペース) に置き換える。
- この件の説明は、「SS-MIX2 標準化ストレージ仕様書 Ver.1.2」 P104 表 3-77 検査結果セグメント (OBX) 定義 の OBX-2 の項目説明にも記述する。

(2) 複数の要素が一つの値で表現されている場合の表記

複数の要素が組み合わせられ一つの結果値として表記されている場合は、それぞれの要素に分離して表記すること。例えば定量値とクラス値が組み合わせられた結果値については、定量値とクラス値に分離する。

【定量値とクラス値の分離の例】

定量値とクラス値が組み合わせられた例

検査名称		院内コード	結果値
ムプス Virus IgG		001591	2.3(±)
↓			
定量値とクラス値を分離した例			
SS-MIX2 標準コード	院内コード	結果値	備考
5F432143102302304	001591	2.3	
5F432143102302311	001591	+-	(半角スペース2つプラスマイナス)

1.3.8 トランザクションストレージのデータ保持期間

トランザクションストレージのデータ保持期間は、現在の標準化ストレージ及び拡張ストレージを作っているデータの再現に必要な分だけ保持しておくこと。

1.3.9 ST 型の長さ

- RXE-23(与薬速度)は ST 型で長さが 6 であるが、正負の記号と小数点を考慮し (例: +266.865)、本事業では 8 桁まで許容するものとする。

- **CX**型は先頭成分が**ST**型で長さが**15**であるが、**IN1-10**(被保険者グループ雇用者 ID)に長い名称の保険者が出力される場合などを考慮し、本事業では**CX**型の先頭成分は**30**桁まで許容するものとする。
- **XAD**型は第**8**成分(その他地理表示)が**ST**型で長さが**50**であるが、全角**50**文字(**100**バイト)と解釈しているシステムがあり半角文字で**100**文字登録出来るため、本事業では**XAD**型の第**8**成分は**100**桁まで許容するものとする。

1.3.10 トランザクションストレージのファイル切り替え機能

SS-MIX2の仕様上、トランザクションストレージはカレントの日付が変わった時点、もしくは記録中のトランザクションデータファイルのファイルサイズが一定量を超えた時点で、新たなファイルを作成して記録先を切り替えるものとなっているが、同一日付内において一定時刻（例えば**17:00**）を経過した時点で記録先を切り替える機能を追加する。

3. 倫理審査における計画書

**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**

研究責任者：堀口 裕正

独立行政法人国立病院機構本部 総合研究センター

診療情報分析部 副部長

事務局/研究主催

独立行政法人国立病院機構本部 総合研究センター

診療情報分析部

堀口 水本

〒152-8621 目黒区東が丘 2 - 5 - 21

TEL: 03-5712-5133

FAX: 03-5712-5134

E-Mail : horiguchi-hiromasa@hosp.go.jp

第 1.0 版 : 2017 年 1 月 18 日

1. 背景

本研究では、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい**退院サマリの自動生成技術の実現**を目的とする。これは用語の標準化を目的とする研究として遠回りな課題設定である。しかし、電子カルテの自動解析は技術的な難易度が高く、実用的な精度を実現するためには多額の研究開発投資が求められる。そこで、本研究提案では、医療現場に直接的なメリットが生じる研究課題に取り組むことによって、現場の協力と今後の追加的な研究開発投資を呼び込み、その過程を通じて実用性の高い電子カルテの自動解析技術を実現する。

2. 目的

本研究は、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい**退院サマリの自動生成技術の実現**を目的とする

3. 研究方法

3-1. 研究実施場所

研究実施場所は、国立病院機構本部総合研究センター診療情報分析部（以下、診療情報分析部）研究室及び本部内分析室並びに静岡大学情報学部行動情報学科狩野研究室、岡山大学大学院医歯薬学総合研究科クリニカルバイオバンクネットワークワーキング事業化研究講座研究室、国立保健医療科学院研究情報支援研究センター研究室とする。

3-2. 研究実施期間

研究実施期間は、倫理審査委員会承認後より 2020 年 3 月 31 日までとする。

3-3. 研究対象医療機関と対象患者

研究対象医療機関は、国立病院機構病院に所属する DPC 病院のうち、診療情報集積基盤（以下、NCDA）を運用しデータ提供を行う医療機関とする。

対象患者は 2016 年 1 月 1 日から 2019 年 12 月 31 日までに入院し、退院時サマリを作成した全患者とする。

3-4. 対象データ

研究に用いるデータは、研究対象医療機関より診療情報分析部に提供されたDPCデータおよびレセプトデータ、ならびにSS-MIX2ストレージに格納された情報から抽出した医師記録、退院サマリおよび入院中の検査結果、食事内容および処方内容である。

3-5. 分析方法

(1) 対象

退院サマリを作成した全患者

(2) アウトカム

入院中に記載/記録された情報から退院サマリを自動生成する技術を開発すること

(3) 抽出する項目

入院中の医師記録・退院サマリ・入院中の検査結果、食事内容および処方内容

(4) 解析方法

入院中に記載/記録された情報を元データに、機械学習により自動的に情報収集を行い、退院サマリを自動で作成する。その作成結果と、実際の医師の書いた退院時サマリを比較/検討を行い、自動作成技術の能力評価を行い、またその能力の改善を行っていく。

4. 倫理的配慮

本研究は、ヘルシンキ宣言、人を対象とする医学系研究に関する倫理指針（以下、倫理指針）に基づいて実施する。

4-1. インフォームド・コンセント

本研究は既存試料・情報を用いて実施し、人体から取得された試料は用いない。研究対象者等からインフォームド・コンセントは受けないが、倫理指針「第12の1 (2) イ」に則り、本計画書の4-3に記す通り、利用目的を含む本研究についての情報を研究対象者等に公開し、研究が実施されることについて研究対象者が拒否できる機会を保障する。なお、NCDA運用による診療情報の蓄積・利活用についての説明及び同意は、各施設での掲示で既に行われている。

4-2. データ管理、個人情報等の取り扱いに関する配慮

研究の実施並びに種々のデータの収集及び取り扱いにおいては、国立病院機構診療情報データベース利活用規程に従うとともに、患者情報の機密保持に充分留意する。

本研究で用いるデータは、研究対象医療機関に2016年1月1日から2019年12月31日までに退院サマリを作成した全患者のデータであり、個人情報等を取り扱う。倫理指針「第15の2(1)」及び国立病院機構診療情報データベース利活用規程に則り、保有する個人情報等について、漏えい、滅失又はき損の防止その他の安全管理のため、下記の措置を講じる。

データは研究対象医療機関で収集され、本部IT推進部に提出される。データが保管されるサーバーを国立病院機構本部2階のセキュリティルームに設置し、セキュリティルーム内でIT推進部システム開発専門職が匿名化処理を行う。研究者は匿名化後のデータを用いて本部内分析室において分析を実施する。

保有する個人情報に関する事項の公表等については、倫理指針「第12の1(2)イ」、「第16の1(1)」及び国立病院機構診療情報データベース利活用規程第6条第3項に則り、個人情報の取扱いを含む研究の実施についての情報を研究対象者等に公開する。

4-3. 本研究における情報公開

本研究では、倫理審査委員会承認後、倫理指針「第12の1(2)イ」、「第16の1(1)」及び国立病院機構診療情報データベース利活用規程第6条第3項に則り、本部ホームページにおいて、本研究の意義、目的及び方法、研究機関、保有する個人情報に関して利用目的の通知、開示、訂正等又は利用停止の求めに応じる手続き並びに保有する個人情報に関する問い合わせや苦情等の窓口の連絡先に関する情報を公開する（公表する情報については別添資料を参照）。

4-4. 研究成果の公表

本研究の成果は、報告書で公表するとともに、学会・論文で発表する。また、本研究結果を内包したソフトウェアの公表を実施する。データの集計・分析結果については、集団を記述した数値データもしくは機械学習の学習結果データとし、個人が同定されるデータの公表は行わない。

5. 研究経費

本研究は、厚生労働科学研究費補助金（臨床研究等ICT基盤構築研究事業）「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」（代表 堀口裕正）を用いて研究を実施する

6. 研究組織

総合研究センター診療情報分析部が主体となり、本部医療部、保険医療科学院、静岡大学、岡山大学等から協力を得て、研究を行う。

【研究代表者】

国立病院機構本部総合研究センター診療情報分析部 副部長 堀口 裕正

【共同研究者】

国立病院機構本部 企画役 岡田 千春

静岡大学情報学部行動情報学科 准教授 狩野 芳伸

岡山大学大学院医歯薬学総合研究科

クリニカルバイオバンクネットワーク

事業化研究講座研究室 准教授 森田 瑞樹

国立保健医療科学院研究情報支援研究センター

特命上席主任研究官 奥村 貴史

別添

「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」研究実施に関するお知らせ