一般口演
# 一般口演9
# 標準化
2024年11月22日(金) 16:30 ～ 18:00 G会場 (504+505)

[2-G-5-02] がん領域における医療機関間連携を目的とした HL7 FHIR mCODEの項目面からの実装の検討

*向井 まさみ[1]、水野　孝昭[2,4]、関 昌佳[5]、柴田　英典[4]、水野　広介[4]、田中 勝弥[3,1]、山本 昇[2]、吉本 世一[1]（1. 国立がん研究センター中央病院医療情報部, 2. 国立がん研究センター中央病院先端医療科, 3. 国立がん研究センター情報統括センター, 4. 株式会社ＣＬＩＮＩＡＬ, 5. 有限会社グローバル・フォー）
*Masami MUKAI[1], Takaaki MIZUNO[2,4], Masayoshi SEKI[5], Hidenori SHIBATA[4], Kosuke MIZUNO[4], Katsuya TANAKA[3,1], Noboru YAMAMOTO[2], Seiichi YOSHIMOTO[1]（1. 国立がん研究センター中央病院医療情報部, 2. 国立がん研究センター中央病院先端医療科, 3. 国立がん研究センター情報統括センター, 4. 株式会社ＣＬＩＮＩＡＬ, 5. 有限会社グローバル・フォー）
キーワード：HL7 FHIR, mCODE, Enterprise Data Base

【背景・目的】当院は急性期がん治療専門病院であり、データ利活用基盤として電子カルテおよび部門システムのテキストデータを集約した統合データベース（以降、統合ＤＢ）を運用中である。統合ＤＢには、当院での実績のデータは存在するが、前医での治療実績等がん領域で治療方針を決める際に必要な情報は診療情報提供書の中に記載されており即座に利用できないという課題が存在する。また、診療情報提供書の HL7 FHIR準拠の記載要領は、2022年厚生労働省標準規格になっているが、記載項目の粒度は従来の診療情報提供書を踏襲しており、疾患の状態や治療歴が即座に利用できるようにはなっていない。今回、がん領域における情報連携の枠組みについて検討したので報告する。【方法】 HL7 FHIRの mCODE（ minimal Common Oncology Data Elements, Version: 4.0.0-ballot）で提案されている項目群に対して、臨床データを設定する項目に対して観察研究、臨床それぞれの目的別に必要度を設定した上で当院の統合 DBに格納されている項目の充足度を確認した。【結果】 mCODEの６つの thematic groups、65FHIR Profileで臨床データから設定する項目は約3400項目であった。そのうち観察研究目的で必要な1076項目のうち292項目（29.4%）、臨床目的で必要な1106項目のうち297項目（29.1%）が統合 DBから即座に抽出可能な項目であった。 Groupごとでは Outcomes および Genomics の充足度が低かった。【考察】 Outcomeの充足度は現状の電子カルテ記事の中に記載されていること、 Genomicsも pdf形式の報告書のみが提供されている状態と矛盾しない。一方、複数の医師からがん領域の情報体系として有用であることも示された。

©一般社団法人 日本医療情報学会

がん領域における医療機関間連携を目的とした HL7 FHIR mCODE の項目面からの実装の検討

向井まさみ[*1]、水野　孝昭[*2*4*1]、関　昌佳[*5]、柴田　英典[*4]、水野　広介[*4]、
田中　勝弥[*3*1]、山本　昇[*2]、吉本　世一[*1]
*1 国立がん研究センター　中央病院　医療情報部、*2 国立がん研究センター　中央病院　先端医療科、
*3 国立がん研究センター　情報統括センター、*4 株式会社CLINIAL、*5 有限会社グローバル・フォー

Consideration of the implementation of HL7 FHIR mCODE for the purpose of collaboration between medical institutions

MUKAI Masami[*1] , MIZUNO Takaaki[*4,3] , SEKI Masayoshi[*5] ,SHIBATA Hidenori[*4] ,MIZUNO Kosuke[*4] ,
TANAKA Katsuya[*3,*1] ,YAMAMOTO Noboru[*2] ,YOSHIMOTO Seiichi[*1]
*1 Division of Medical Informatics, National Cancer Center Hospital,
*2 Department of Experimental Therapeutics, National Cancer Center Hospital
*3 IT Integrating and Support Center, National Cancer Center,
*4 CLINIAL CO., LTD, *5 GLOBAL-For corp.

The results of the examination of necessary items for each purpose of use for the mCODE proposed by HL7 as a necessary data set for the oncology field were about 60%. The highest number of items that can be set immediately from the data stored by the current system was 81%, which can be provided from the integrated DB, while the lowest percentage was 0%. In order to increase the number of items that can be extracted from the system, extensions such as electronic linkage of test results from other institutions and the ability to extract structured items using generative AI technology were thought to have the potential to make data management more efficient. On the other hand, continued verification of the comprehensiveness of mCODE was considered necessary.

Keywords: HL7 FHIR, mCODE, Enterprise Database, DWH.

1. 背景・目的

1.1 現状と課題

近年、診療データを源としたリアルワールドデータの利用による AI 開発や治験や創薬等の臨床研究分野におけるリアルワールドエビデンスの創出が注目されている。当院は急性期がん治療専門病院であり、ほぼ全ての患者が他院から紹介され、当院で診断、治療を受けている。がん治療においては初発時の診断やそれに伴う検査結果、その後の治療歴および投薬情報等は当院での治療方針を決めるための重要な情報である。現状の診療での医療機関間の情報共有は、文書情報として、紹介状（診療情報提供書）の中に記載されており、当院では紙媒体で受領した文書をスキャンしたうえで電子カルテ端末情報で参照する仕組みとなっており、即座に記載情報本体を電子的に利用できる状態には至っていない。また、標準規格の視点では、診療情報提供書の HL7 FHIR 準拠の記載要領は、2022 年厚生労働省標準規格になっているが、記載項目の粒度は従来の診療情報提供書を踏襲しており、疾患の状態や治療歴が即座に利用できるようにはなっていない。

1.2 検討を開始するに至った背景

1.2.1 HL7 mCODE の開発

がん領域では、2019 年に HL7 International により mCODE™（Minimal Common Oncology Data Elements）[1]が公開された、mCODE は、電子カルテ等で利用可能ながん領域/腫瘍学における構造化されたデータ要素のコア・セットを組み立てることを目的とした取り組みである。mCODE のデータ構造は HL7 FHIR を利用して検討されているため、本邦における厚生労働省標準規格となった3文書6情報の規格との親和性が高いと考えられる。

1.2.2 当院のデータ利活用基盤の構築

診療データを活用するには、その収集、加工および抽出の仕組みが必要となる。院内で稼動する情報システムは、オーダエントリ機能をもつ電子カルテシステムをはじめ、放射線、生化学検査、生理検査、薬剤、手術など約 60 の部門システムから構成されている。我々は、これらの分散管理されている診療データを集約した診療用統合データベースを 2021 年に構築した（図1）。[2] [3] [4]

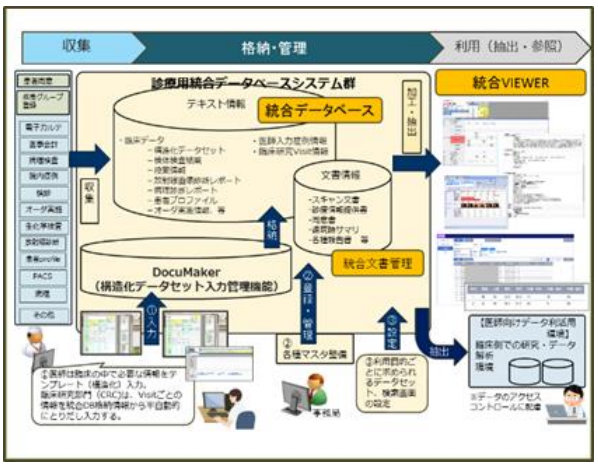

図1. データ利活用基盤”統合データベースシステム“概要

統合データベースには稼働システムのデータほかに、従来診療科や部門が個別で運用するデータも移行集約している。収集したデータは、その種別によりアクセスコントロールを付与したうえで、一次利用、二次利用できる環境を構築し運用されている。

1.3 目的

統合DBには、当院での実績のデータは存在するが、前医での診断や治療実績等がん領域で治療方針を決める際に必要な情報は診療情報提供書の中に記載されており即座に利用できないという課題が存在する。先述した mCODE 形式で記載されたデータセットが前医からの診療情報提供書の補足情報のような形で連携すること、また当院での診療を終了したあとの紹介先施設への連携情報として当院での実績情報を含めた診療データが生成できれば、スキャンデータ等からの転記作業によるデータ品質の精度向上や業務負担を軽減できることが期待できる。

今回、がん領域における情報連携の枠組みとして mCODE の活用について検討したので報告する。

2. 方法

HL7 FHIR の mCODE(minimal Common Oncology Data Elements, Version: 4.0.0-ballot)で提案されている項目群に対して、臨床データを設定する項目を整理した上で観察研究、治験、臨床それぞれの目的別に必要な項目を抽出し、当院で稼働しているシステムごとに格納されている項目の充足度を確認した。

2.1 mCODE で定義されている項目の確認

2.1.1 検討対象項目の精査

標準規格が決めるデータ形式には、連携するシステムが解釈および処理を行うための項目が多く含まれている。本研究では、臨床で発生するデータ項目について検討を行うことが目的のため、システム処理のために必要な項目は除外する作業を行った。除外した基準を以下に示す。

a. リソース ID などのようにシステム(ソフトウェア)がデータを生成するときに機械的に割り当てるもの
b. 詳細項目が値を持っている場合の、その上位項目(中／大項目) ∵項目数が重複してしまうため
c. 氏名、電話番号、住所など構造化のために細分化された項目で日本の情報システムでは構造化項目を使用していないことが明らかなもの。
d. 他のリソース等を参照している項目
e. 定義上予約されているが、現実的に値を保持する可能性がないもの
f. プロファイルとして使用しないと明記されている項目
g. Bundle

2.1.2 ユースケースごとに必要な項目の精査

診療データはそれを利用するシーンにより、必要な項目が異なる。例えば、後方的観察研究においては患者氏名や詳細な居住住所情報等の必要性は高くないが、臨床現場で同意書に署名をいただく場合は医療安全上患者本人であることを確認するための情報は必須となる。そこで、本研究では、ユースケースとして観察研究、治験、臨床の3つのユースケースを想定し、それぞれのシーンで必要となる項目を決定した。項目の決定は、臨床および研究の経験がある医師がすべての項目を確認しその要否を判断した。そのうえで、診療情報管理師の資格を持つものが、レビュを行うことで確定した。

2.2 保有するシステムごとの項目充足度の算出

現在当院が提供できるデータソースは、電子カルテが提供する DWH、先述した統合データベースおよび診療科がそれぞれ入力している構造化テンプレートの3種類が存在する。そこで、このデータソースごとに保有するデータ項目から mCODE の項目に設定可能なものを確認した。抽出確認にあたっては、一部簡単な変換を含むが、ほぼ即座に項目が利用できるものを「利用可」として定義した。また、構造化テンプレートについては外科系、内科系の診療科のテンプレートを評価した。診療科別に治療方法等が異なるためいずれかのテンプレートにあるものは「利用可」と定義した。

これらを先に整理したユースケースごとに必要な項目を母数として、設定可能な項目の割合を算出し充足度とした。

3. 結果

検討対象とした mCODE は、Patient, Disease, Genomics, Assessment, Treatment, Outcome の 6 つの thematic groups、65 の FHIR Profile で構成されている。 本研究では、このうち複数の thematic group から参照されている要素は、General group として独立させて検討した。mCODE の全項目のうち、方法1で設定した除外条件を反映した臨床データから設定する項目は 3,326 項目であった。Thematic Group ごとの利用目的別の必要な項目数と3つのシステムから即座に抽出できる項目数の一覧を表1に、利用目的別の各システムから抽出可能な項目数の割合を表2に示す。

表2. Thematic Group ごとの
利用目的別のシステムから抽出可能な項目数に割合

Thematic Group	利用目的	DWH	統合DB	診療科別テンプレート
Patient	臨床研究	42.3%	80.8%	88.5%
	治験	42.3%	80.8%	88.5%
	臨床	35.5%	67.7%	74.2%
Disease	臨床研究	1.0%	12.4%	21.0%
	治験	1.0%	12.4%	21.0%
	臨床	1.0%	12.4%	21.0%
Genomics	臨床研究	0.0%	0.0%	22.1%
	治験	0.0%	0.0%	22.2%
	臨床	0.0%	0.0%	22.1%
Assessment	臨床研究	11.2%	17.1%	26.5%
	治験	11.2%	17.1%	26.5%
	臨床	11.2%	17.1%	26.5%
Treatment	臨床研究	34.7%	55.3%	72.0%
	治験	35.6%	56.8%	74.0%
	臨床	31.1%	49.7%	64.7%
Outcome	臨床研究	0.0%	0.0%	72.2%
	治験	0.0%	0.0%	72.2%
	臨床	0.0%	0.0%	68.4%
Genenal	臨床研究	0.0%	0.0%	0.0%
	治験	0.0%	0.0%	0.0%
	臨床	0.0%	0.0%	0.0%
総計	臨床研究	4.7%	14.9%	27.6%
	治験	4.8%	15.0%	27.7%
	臨床	4.7%	14.7%	27.2%

利用目的にかかわらず充足度（抽出可能な項目数／必要な項目数）は診療科別テンプレート、統合データベース、電子カルテ DWH の順で高い値を示した。もっとも充足度が高いのは Patient Profile の診療科別テンプレートで 88.5%を示した。一方で、Genomics, Outcome の情報は統合 DB, DWH いずれも 0%であった、

4. 考察

本研究でとりあげた3つのユースケースごとに必要な項目数に大きな差異はなかった。選択した項目の種類では、臨床では必要とされるが治験では必要ではない項目の差異は存在した。thematic group 別で充足度が低い Genomics は、現在はゲノム検査設備が院内には存在せず外部機関の結果情報が pdf 等で提供される運用であること、Outcome は治療の予後情報や奏功にかかわる項目群であるが電子カルテの記事記載やフォローアップのための診断レポートには記載されているが即座に抽出できない現状と矛盾しない。

システム別の充足度は収集している情報が多岐にわたる統合 DB のほうが DWH よりもやや良い結果となった。診療科テンプレートの情報はほぼ全項目が構造化されたデータであること、原発、転移巣についての情報、TNM 分類、がんのリスクになった喫煙や飲酒などの因子、病理に関する情報、治療歴に関する情報も記録されているため、全体を通してもっとも充足度が高い結果となった。これは、がん種ごとの差異はあるが、テンプレートが疾患レジストリ登録や NCD に登録するために記載されているため、必要な項目の網羅性が高いことが考えられる。また、現在では治療方針を決めるために Genomics の情報は重要なため、テンプレートでは pdf の結果が転記されていることが想定された。

今後、システムから抽出する項目の充足度を上げるためには、生成 AI や LLM (Large Language Model)の技術を利用して普段記載されている SOAP 形式の記録や診断レポートから必要な項目を抽出する機能を実装する等の工夫が考えられる。

今回は、mCODE に対して院内システムから設定可能な項目群を評価したが、実際に mCODE 形式のデータセットを生成するためには、SNOMED-CT のように mCODE で参照されている標準規格コードが日本では即座に利用できない等の課題もあることがわかった。これについては、厚生労働省標準規格に置き換える、National Extension として申請を進める等いくつかの対策を考える必要がある。

がん領域の専門医からは、連携すべき情報として初見の範囲で有用であるとのコメントを得たが、客観的な評価には至っていない。本件は、今後日本でのがん種や治療別の疾患レジストリ項目に対する網羅性を確認していく必要があると考えている。

5. 結語

がん領域に必要なデータセットとして、HL7 が提案する mCODE が定義する項目群に対して、データの利用目的ごとに必要な項目を検討した結果は、全体の約 60%程度であった。その項目群に対して現状のシステムが保管しているデータから即座に設定できる項目はもっとも高いもので統合 DB から提供できる 81%であったが、もっとも低い割合は 0%であった。システムから抽出できる項目を増やす対策として、他機関の検査結果の電子的な連携、生成 AI 技術等を利用した構造化された項目の抽出機能などの拡張を行い、データ収集および管理を効率的にすることが挙げられる。一方で、mCODE の網羅性について継続検証が必要であると考えられる。

6. 謝辞

本研究は、国立高度専門医療研究センター医療研究連携推進本部横断的研究推進費（課題番号：2022-B-04）の支援によって実施されている。ここに記載し感謝申し上げます。

参考文献

1) mCODE [https://build.fhir.org/ig/HL7/fhir-mCODE-ig/ (cited 2024-Aug-28)]
2) 1) 向井まさみ,田中勝弥,,三原直樹, 診療情報利活用のための SS-MIX2 ストレージを対象としたデータバリデーション手法の提案,第 42 回日本医療情報学連合大会,2022
2) 向井まさみ,田中勝弥,三原直樹,浜本隆二. 診療情報の研究的二次利用を目的とした統合的な診療情報利活用基盤の開発,第 3 回日本メディカル AI 学会学術集計,2021.
3) 三原直樹,向井まさみ、他. 統合参照ビューアの診療科医師による評価とその解析,第 40 回日本医療情報学連合大会,2020

表1. 利用目的ごとに必要な項目とシステムごとに抽出可能な項目数一覧

Thematic Group	全項目数	利用目的ごとに必要な項目			データソースごとに抽出可能な項目		
		臨床研究	治験	臨床	DWH	統合DB	診療科別テンプレート
Patient	516	26	26	31	11	21	23
Disease	1,439	1,363	1,361	1,365	14	169	286
Genomics	508	263	261	263	0	0	58
Assessment	368	170	170	170	19	29	45
Treatment	334	150	146	167	52	83	108
Outcome	156	54	54	57	0	0	39
General	5	1	1	1	0	0	0
合計	3,321	2,027	2,019	2,054	96	302	559

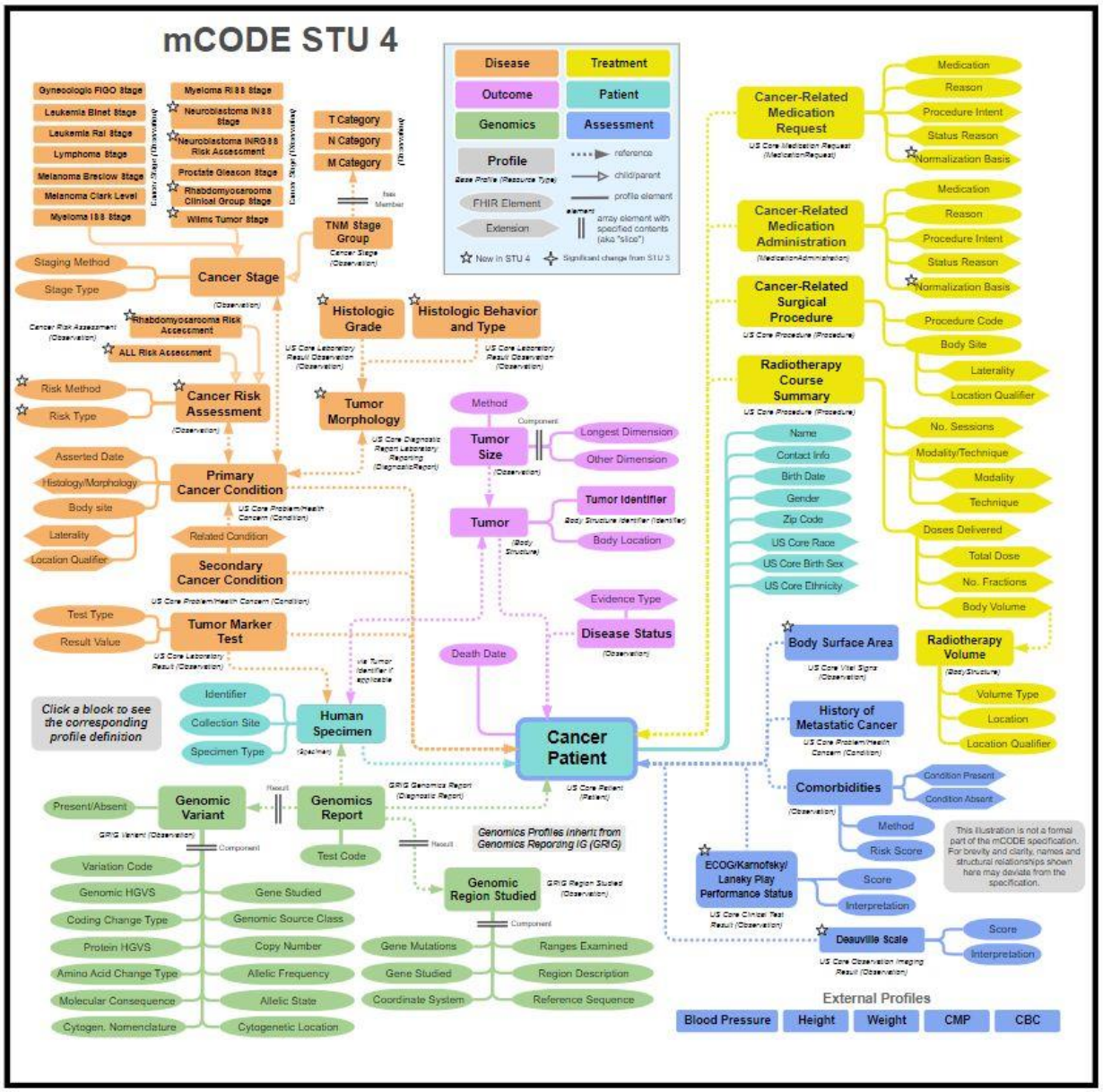

図2.　HL7 mCODE の構造][1]