

一般口演

一般口演24

DWH・CDISC・その他

2017年11月22日(水) 16:00 ～ 17:30 F会場 (10F 会議室1004-1005)

[3-F-3-OP24-4] 造血器腫瘍領域における臨床ゲノム情報データストレージの開発における CDISC標準の利活用

山本 松雄, 近藤 修平, 堀部 敬三, 齋藤 俊樹（国立病院機構名古屋医療センター）

【背景】

次世代シーケンス由来のゲノムデータは情報量が多く、通常の臨床検査結果と比較すると診断や適切な治療内容など、結果の臨床的意義を体系化・カテゴライズするのが圧倒的に困難である。そのため「ゲノム情報」と「臨床情報」が紐付いた状態で大量に蓄積されたデータベースに検査した患者さんのゲノムデータ等を照合し、臨床的意義を導き出すことが必要である。一方、複数の臨床試験由来のデータをまとめて検索・抽出するには臨床情報の標準化が必要である。

【目的】

詳細なゲノム情報とが紐付いた状態で格納できるデータベースを構築する上で、臨床情報を CDISC標準に変換する

【方法】

CDISC SDTM形式を利用し、SDTM実装ガイド(3.2版)に則ったマッピングを行う。

【結果】

2試験、約800症例分のデータをマッピングしたところ、SDTM実装ガイドでは不十分でより詳細な仕様が必要であった。仕様書を作成し、以下に公開した。

https://www.nnh.go.jp/crc/departments/information_system_research/

また問題点として次のことが明らかになった。

再発イベントは、その試験のプロトコールに依存して、DSドメイン(Disposition)やCEドメイン(Clinical Events)に格納されるが、データをまとめて検索・抽出するにはいずれかの単一ドメインにマッピングする必要がある。データベースでは、CEドメインへマッピングすることを仕様とした。

すでに CDISC標準を利用せずに完了した試験については、CDISC SDTM形式へのマッピングが必要となり、多くのリソースを費やした。また蓄積されたデータベースのBIツールを利用したデータの可視化や即時的な解析を可能とした。

造血器腫瘍領域における臨床ゲノム情報データストレージの開発における CDISC 標準の利活用

山本 松雄^{*1}、近藤 修平^{*1}、堀部 敬三^{*1}、齋藤 俊樹^{*1}

^{*1} 国立病院機構名古屋医療センター

Use case of CDISC standards on development of the clinical-genomic database for hemato-oncology

Matsuo Yamamoto^{*1}, Shuhei Kondo^{*1}, Keizo Horibe^{*1}, Toshiki Saito^{*1}

^{*1} National Hospital Organization Nagoya Medical Center

Abstract:

Genome data derived from the next generation sequence has a large amount of information. The genomic data is very difficult to systematize and categorize the clinical significance of the results as compared to the usual laboratory test results. Therefore, it is necessary to accumulate a large amount of "genomic information" and "clinical information" in the database in a state of being linked, and to derive the clinical significance by comparing the genomic data etc. of the examined patient to the database. On the other hand, standardization of clinical information is necessary to retrieve and extract data derived from multiple clinical trials. For standardization, we use the CDISC SDTM format and map clinical information according to the SDTM Implementation Guide (3.2 version). We mapped the data for 2 studies and about 800 cases, and it turned out that the SDTM implementation guide was insufficient and it was found that more detailed specifications were necessary, so we made a specification sheet. Several problems have also become apparent. Studies that were completed without using the CDISC standard required mapping to CDISC SDTM format and spent a lot of resources.

Keywords: Genome, database, hematological malignancies, CDISC, SDTM

1. 緒論

次世代シーケンス由来のゲノムデータは情報量が多く、通常の臨床検査結果と比較すると診断や適切な治療内容など、結果の臨床的意義を体系化・カテゴライズするのが圧倒的に困難である。そのため「ゲノム情報」と「臨床情報」が紐付いた状態で大量に蓄積されたデータベースに検査した患者さんのゲノムデータ等を照合し、臨床的意義を導き出すことが必要である。一方、複数の臨床試験由来のデータをまとめて検索・抽出するには臨床情報の標準化が必要である。

平成 26 年 6 月 20 日に厚生労働省より「承認申請時の電子データ提出に関する基本的考え方」が発出され、承認申請の対象となる臨床情報については、平成 28 年 10 月 1 日より(経過措置期間は平成 32 年 3 月 31 日まで)Clinical Data Interchange Standards Consortium の規格(以下「CDISC 標準」という。)に準拠した形式での電子データの提出が開始した。米国、欧州、中国においても臨床情報の標準として、CDISC 標準を採用し、CDISC 標準が臨床情報の国際標準となっている。

2. 目的

詳細なゲノム情報と臨床情報が紐付いた状態で格納できるデータベースを構築する上で、臨床情報を CDISC 標準に変換する。

3. 方法

CDISC 標準の一つである SDTM 形式(1.4 版)¹⁾を利用し、SDTM 実装ガイド(3.2 版)²⁾に則ったマッピングを行う。カテゴリ値等の統制用語については、SDTM Terminology(2016-12-16 版)³⁾に従う。マッピングの対象のデータは、臨床試験由来のデータのうち以下に限定した。

1. 背景情報(登録日、生年月日、年齢、性別、群(割付)、国、初回投与日、同意取得日)
2. 再発(再発確認日)

3. 2 次がん(2 次がん確認日)
4. 中止・完了(中止理由(死亡を含む)、中止日、完了日(最終生存確認日))
5. 疾患情報(原疾患名、診断日)
6. リスク情報(リスク分類)
7. 寛解情報(効果判定)

4. 結果

2 試験、約 800 症例分のデータをマッピングしたところ、SDTM 実装ガイドでは不十分で、より詳細な仕様が必要であった。仕様書を作成し、以下に公開した。

https://www.nnh.go.jp/crc/departments/information_system_research/

また問題点として次のことが明らかになった。

SDTM 形式へのマッピングは、その試験のプロトコールに依存しており、再発イベントが中止条件である場合は、DS ドメイン(Disposition Events)に格納されることが考えられ、中止条件でない場合は、CE ドメイン(Clinical Events)に格納されることが考えられる。しかしながら、データをまとめて検索・抽出するにはいずれかの単一ドメインにマッピングする必要があり、臨床ゲノムデータストレージでは、CE ドメインへマッピングすることを仕様とした。

疾患名のコーディングに使用する辞書について、造血器腫瘍領域においては、WHO 分類にてコーディングされることがニーズとしてある。しかしながら、WHO 分類の版は、下位互換性がなく、分類に必要な情報も版ごとに異なるため、常に最新の WHO 分類でコーディングすることができない。臨床ゲノムデータストレージでは、ICD10 対応標準病名マスターにてコーディングすることを仕様とした。

すでに CDISC 標準を利用せずに完了した試験については、SDTM 形式へのマッピングが必要となり、多くのリソースを費やした。また蓄積されたデータベースのデータの可視化や即時的な解析を可能とした。

5. 考察

臨床ゲノムデータストレージが知識ベースとして利用されるまで成熟した際に、その利用方法が、利用者に依存せず固定化しているのであれば、データの可視化や即時的な解析のバリエーションは固定化される。しかしながら、利用者に依存して利用方法は異なる。その解決策として、データの可視化や即時的な解析をフレキシブルにすることが可能である BI ツールの利用が有用であると考ええる。

このデータベースに蓄積される臨床情報は、基本的な背景情報および予後に関わる情報に限定しているが、将来的に、このデータベースが巨大化した際には、この知識ベースは「人間が読み取り可能な知識ベース」から「機械 (AI) が読み取り可能な知識ベース」になることが考えられる。その際には、情報は多ければ多いほど良く、情報として有用であるかどうかはあまり判断せず、可能な限り多くのデータを蓄積することも検討しなければならない。

6. 結論

造血器腫瘍領域における臨床ゲノム情報データストレージの開発にて、臨床情報を SDTM 形式へマッピングすることにより、複数の臨床試験由来のデータをまとめて検索・抽出することを可能にした。

参考文献

- 1) CDISC Submission Data Standards Team. Study Data Tabulation Model Version 1.4. Clinical Data Interchange Standards Consortium, 2013.
[https://www.cdisc.org/system/files/members/standard/foundational/sdtm/study_data_tabulation_model_v1_4.pdf (accessed 2017-Sep-7)].
- 2) CDISC Submission Data Standards Team. Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.2. Clinical Data Interchange Standards Consortium, 2012.
[https://www.cdisc.org/system/files/members/standard/foundational/sdtmig/sdtmig_20v3.2_20noportfolio.pdf (accessed 2017-Sep-7)].
- 3) National Cancer Institute. CDISC SDTM Controlled Terminology, 2016.
[<https://evs.nci.nih.gov/ftp1/CDISC/SDTM/Archive/SDTM%20Terminology%202016-12-16.pdf> (accessed 2017-Sep-7)].