

課題番号 : 26指114
 研究課題名 : 臨床データのCDISC標準化に関するフィージビリティスタディ
 主任研究者名 : 大津 洋 (2015年までは田中康博)
 分担研究者名 : - (2014, 佐藤泉/ 2014-2015, 川崎洋平)
 キーワード : CIDSC, バイオバンク, 臨床データ, データマネジメント, 標準化
 研究成果 :

研究概要

NCGM はじめ NC が保有するバイオバンクのデータ（臨床データや検体データ）は、臨床研究（試験）や治験に利用される。データが治験に利用されるためには、主に海外製薬企業が中心となり FDA への電子申請文書の電子データ形式として採用され、普及してきた CDISC 標準に対応できることも必要であると考えられる。

一方、JCRAC データセンターでは主に多施設の研究者（医師）主導の臨床研究データマネジメントを請け負っているが、NCGM が行う医師主導治験のデータセンターを担えるためには、PMDA へ提出する電子申請文書が 2016 年から CDISC 形式による受け入れ義務化を控え、医師主導治験が実施された暁には、十分に対応できることが必要となる。

そのため、バイオバンクと臨床研究データマネジメントの両局面において共通する臨床データについての CDISC 標準化のための技術開発と実証評価を行う。その一方で、標準化後の臨床データの利用についても可能な限り国際的な動向を調査する。

方法・研究結果

Biobank:

(Step1) CDISC SDTM とバイオバンクのデータフォーマットのマッピングを行う

(Step2) CDISC ADaM に変換可能かどうかを検討する。

JCRAC(Clinical Data):

(Step1) 基幹システムとして、REDCap を導入する・GAP 分析を行う

(Step2) REDCap 上で、CDISC の一連のデータ交換・分析が可能かを調査・研究を行う。

	2014	2015	2016
Biobank	← Step 1 →	← Step 2 →	
JCRAC (Clinical Data)	← Step 1 →		← Step 2 →

初年度は国内外の CDISC 関係の動向を生で把握するため、CDISC Japan ユーザー会に参加し、連携を図った。CDSIC の中心となる SDTM (Study Data Tabulation Model: 申請臨床試験データモデル) のユーザー会活動を始め、企業、アカデミアのユーザー会メンバとの情報交流を行った。一方、通信規約 ODM (Operational Data Model) については、調査研究の結果拠点間の連携を行うためには、開発費には当時、少なくとも 3000 万円程度はかかり、3~5 年以上かけてじっくり取り組まないと難しいと判断した。ODM については基本検討や期間内での実例調査にとどめ、継続的に調査を進めるものの、核となる SDTM に注力した取組みとすることとした。

(JCRAC データセンター)

本研究を進める上での研究環境として EDC (Electronic Data Capture) の REDCap を導入し、サーバーを 2014 年 3 月に稼動を開始した。本研究の狙いは REDCap を使って、これに CDISC 対応の実証評価を

行うことである。まず REDCap 自身の使い方、利用習熟に取り掛かった。REDCap は Vanderbilt 大学の開発であるため、操作説明、マニュアルに相当するものが日本語化されていない状況下で、JCRAC データセンターで受託している臨床研究の症例登録票 (CRF) を試験的に利用し、症例登録画面の作成や、データのアップロード、ダウンロードの等、基本的な機能の確認をとおして、REDCap 機能を理解し、本システムの適用範囲 (限界) を調べた。2015 年 3 月末時点で、システムの適正な動作運転がなされていると判断し、JCRAC の基幹システムとして公開した。



2-3 年度目は、REDCap 上での CDISC 評価のため、JCRAC で模擬臨床研究を立ち上げて連携を行うこととした。当初は SDTM ユーザー会と連携して、アカデミアと企業が同じシステム上にデータを入れることを想定し準備を進めていたが、十分な期間が取れないことにより、内部プロジェクト評価を行うこととした (

左図)。同時に、REDCap 自身も CDISC 対応するモジュールを公開したため、JCRAC もバージョンアップし、CDISC が基本としている ODM フォーマットの

入出力が可能となったため、期間内でのフィージビリティ研究が容易となった。当初想定していた計画としては違うが、結果的には予定通りのフィージビリティ研究が実施でき、日本臨床試験学会学術総会にてその成果を発表できた。

この中で、臨床データの交換には、CDISC ODM を使うことが可能であること、ODM 形式であれば、解析ソフトウェアの SAS などの CDISC と親和性の高いものであれば、バージョンの問題に注意すれば利用は容易であることを発表した。命名規約については、CDISC の規約 (CDASH など) を理解しておく必要性はあるが、これは、テンプレートを今後整備していけばよいと判断している。

最近の日本のアカデミアの発表の場合、CDISC の情報規約である ODM を無視した規格の利用についての発表が多いが、我々は CDISC の基本に忠実に対応が可能であることを明言しているのが特徴的である。これにより、CDISC だけではなく、医療情報の標準規格 (HL7 Version3/ XML フォーマット) といった XML 記述のものに対しても応用可能になる基盤ができたと考えている (下図)

Input example		Context
Type1	Using SS-MIX Data	SS-MIX data because it has become a basic CSV, character type and the project format If the note only possible incorporation. But, in the HL7 format of the format it was necessary to remake all.
Type2	Using CDISC ODM	ODM format from the HIS could not be output. It attempts to create according to the REDCap of ODM format Although uptake is an error. In REDCap Ver6.10.3 it was not able to incorporation.
Type3	Using CSV_File	no problem. If the note only character type and the project of the format was possible incorporation.
Output Example		
TypeA	Using CDISC ODM/SDTM	ODM format from the HIS could not be output. It attempts to create according to the REDCap of ODM format Although uptake is an error.
TypeB	Using (SAS/SPSS)CSV etc.	no problem.

なお、標準化の実証評価の途中で、IMDRF (International Medical Device Regulatory Forum) やレギュラトリーサイエンスなど国際協議機関に参加でき、実証後の利活用について議論が可能となった。実証実験の先の利活用を検討するに値する CDISC 等への知見や、標準化の意識の高まりを受け、IMDRF

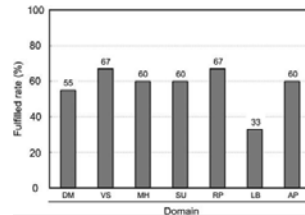
methodology WG にPMDA のミラーグループ一員(non regulatory agency)として、London(UK)/Moscow(Russia) F2F meeting の参加、また、その後の Web meeting への参加を通じて、臨床試験だけではなく、臨床疫学やレジストリ研究におけるデータの標準化について議論した。その成果の一部は、IMDRF document などの作成に貢献した。

(バイオバンク)

分担研究者の佐藤(当時)を中心に、NCBN 中央データベース管理部門でカタログデータベースに用いている個票データフォーマットの項目をSDTM のドメイン、変数にどこまで変換が可能かについて調べた。左図は、NCBN 中央管理部門が持っている個票データフォーマットとのマッピングモデル、右側がそれを基とした、適合割合を支援した。

The screenshot shows a software interface for mapping data items. It is divided into two main sections: '1. Classification of items' on the left and '2. Determine the domain' on the right. The left section lists various data items with their categories and codes. The right section shows a table of domains and their corresponding codes. An arrow points from the left section to the right section, indicating the mapping process.

おおよそ全体の7割程度であるが、治験ではバイオバンク独自の検体に関してSDTM用に変数を定義することが分かった(図)。



度の項目は変換可
使用しないバイオ
する項目など、独自
する必要があるこ

その結果については、日本計算機統計学会第28回シンポジウム(2014年11月14日、15日、沖縄科学技術大学院大学)および日本臨床試験学会第6回学術集会総会(2015年2月20日)で発表し、論文としてまとめ発表された。

2年度目は、分担研究者の川崎(当時)を中心に、佐藤の実施したSDTM変換のフィージビリティスタディの結果を受けて、ADaM(Analysis Dataset Model:申請統計解析データモデル)への変換についての調査を行った。その結果については、日本臨床薬理学会第36回学術総会(2015年12月9日~11日)にて発表した。バイオバンクデータのCDISC化についてのフィージビリティスタディの基本的な検討は目標を達成した。

現時点の調査した限りでは、バイオバンクデータに関する国際的な標準化の目立った動きはなく、ましてやCDISCというキーワードが飛び交う状況ではないという判断のもと、臨床試験や治験を実施する上で、バイオバンク試料等を利用するという局面が出てくるのであれば、むしろそのような展開になるべきと考えるが、それを見越した地ならしを進めておき、業界のデータ標準化を先導的に牽引していくという視点では、取り組むべき価値は十分あると考える。

Subject No. : 26A-114
Title : Feasibility study on CDISC standards in Clinical data
Researchers : Hiroshi Ohtsu (2015-2016)
Yasuhiro Tanaka (2014-2015)
Kohei Kawasaki (2014 -2015)
Izumi Sato (2014)
Key word : CDISC, Biobank, clinical data , data management, scandalization
Abstract :

Biobank data (clinical data and specimen data) owned by NCGM, including NCGM, are used for clinical research (trials) and clinical trials. In order for data to be used for clinical trials, it is necessary to be able to respond to the popular CDISC standard, mainly adopted as an electronic data format of electronic application documents to FDA, mainly by overseas pharmaceutical companies.

Otherwise, JCRAC data center is mainly undertaking clinical research data management led by multicenter researchers (physicians), but in order to be able to bear the data center of doctor-initiated clinical trial conducted by NCGM, the electronic application document to be submitted to PMDA is obliged to accept CDISC format from 2016, and it is necessary to be able to respond adequately when the doctor-led clinical trial is implemented.

For this reason, we will conduct technical development and empirical evaluation for CDISC standardization on common clinical data in both phases of biobank and clinical research data management. And, we will investigate international trends as much as possible about the use of standardized clinical data.

Method:

We designed the feasibility plan for 2 types data as follow:

• Biobank:

1. Mapping between CDISC SDTM domains / variable and individual data format using in NCBN central database management
2. From result of #1 , we examined the applicability to CDISC ADaM

• JCRAC(Clinical data):

1. Adopt REDCap (Vanderbilt University) as JCRAC's backbone system to use data management and use CDISC standards
2. Feasibility Study using REDCap (EDC -> SDTM or ODM -> ADaM)

Researchers には、分担研究者を記載する。

	2014	2015	2016
Biobank	Step 1		Step2
JCRAC (Clinical Data)	Step1	Step2	

Result:

(Biobank)

As we reported in I.Sato et al (2016), regarding the problematic but classifiable items, out of the variables of medical history, surgical history, transfusion history, and family history, the reasons for the problematic classifications were because of disparities in the methods of categorizing the selections and the treatment of date data from the SDTM. In addition, some question items and existing variables had to be changed and supplemental materials created. A complete breakdown of the “problematic but classifiable” items is presented. Overall, 50%–70% of the required items in each domain were fulfilled.

Common additional items such as STUDYID, USUBJID, and ARM needed to be included to achieve 100.0% fulfillment of the required items. The proportion of variables in the LB domain was low at 33% because additional test items such as LBTESTCD and LBTEST needed to be included (Fig 1).

Of course, there is possible also converted to ADaM (Kawasaki et al,2015), but, it is necessary to pay attention to the SDTM compatibility of each domain.

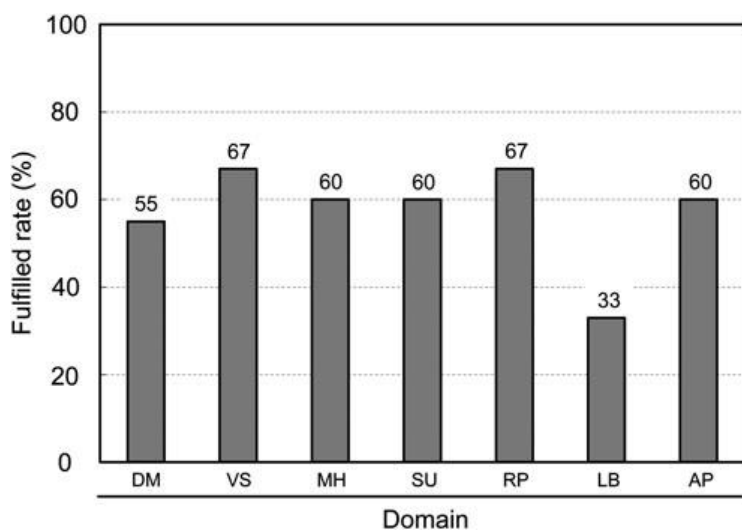


Figure 1 Proportions of required items that were fulfilled by the National Center Biobank Network data in each domain (%).

(JCRAC)

We succeeded in introducing REDCap as a core data management system for clinical research, and we upgraded to REDCap which corresponded to CDISC ODM format. (Step 1)

At the same time, because discussion was also being made on research utilization of medical information, we conducted a feasibility study as an object of survey here as well.

Input example		Context
Type1	Using SS-MIX Data	SS-MIX data because it has become a basic CSV, character type and the project format of If the note only possible incorporation. But, in the HL7 format of the format it was necessary to remake all.
Type2	Using CDISC ODM	ODM format from the HIS could not be output. It attempts to create according to the REDCap of ODM format Although uptake is an error. In REDCap Ver6.10.3 it was not able to incorporation.
Type3	Using CSV_File	no problem. If the note only character type and the project of the format was possible incorporation.
Output Example		
TypeA	Using CDISC ODM/SDTM	ODM format from the HIS could not be output. It attempts to create according to the REDCap of ODM format Although uptake is an error.
TypeB	Using (SAS/SPSS)CSV etc..	no problem.

It is distinctive that we clearly state that it is possible to faithfully respond to the basic of CDISC. By doing this, we believe that the infrastructure has become available not only for CDISC but also for XML description such as medical information standard (HL 7 Version 3 / XML format).

Researchers には、分担研究者を記載する。

研究課題名 : 26指114 臨床データのCDISC標準化に関するフィージビリティスタディ
主任研究者名 : 大津 洋
分担研究者名 : -
キーワード : CIDSC, バイオバンク, 臨床データ, データマネジメント, 標準

研究背景

NCGMはじめNCが保有するバイオバンクのデータ（臨床データや検体データ）は、臨床研究（試験）や治験に利用される。データが治験に利用されるためには、主に海外製薬企業を中心となりFDAへの電子申請文書の電子データ形式として採用され、普及してきたCDISC標準に対応できることも必要であると考えられる。一方、JCRACデータセンターでは主に多施設の研究者（医師）主導の臨床研究データマネジメントを請け負っているが、NCGMが行う医師主導治験のデータセンターを担えるためには、PMDAへ提出する電子申請文書が2016年からCDISC形式による受け入れ義務化を控え、医師主導治験が実施された暁には、十分に対応できることが必要となる。そのため、バイオバンクと臨床研究データマネジメントの両局面において共通する臨床データについてのCDISC標準化のための技術開発と実証評価を行う。その一方で、標準化後の臨床データの利用についても可能な限り国際的な動向を調査する。

研究課題名 : 26指114 臨床データのCDISC標準化に関するフィージビリティスタディ
 主任研究者名 : 大津 洋
 分担研究者名 : -
 キーワード : CIDSC, バイオバンク, 臨床データ, データマネージメント, 標準

研究方法

本研究が扱う情報は、中央バイオバンクが持つカタログデータ（以下、Biobankデータ）とJCRACデータセンターが取り扱う臨床研究データ（以下、JCRAC）がある。それぞれをStep1,2とし、期間内でのフィージビリティ研究を行うこととした。



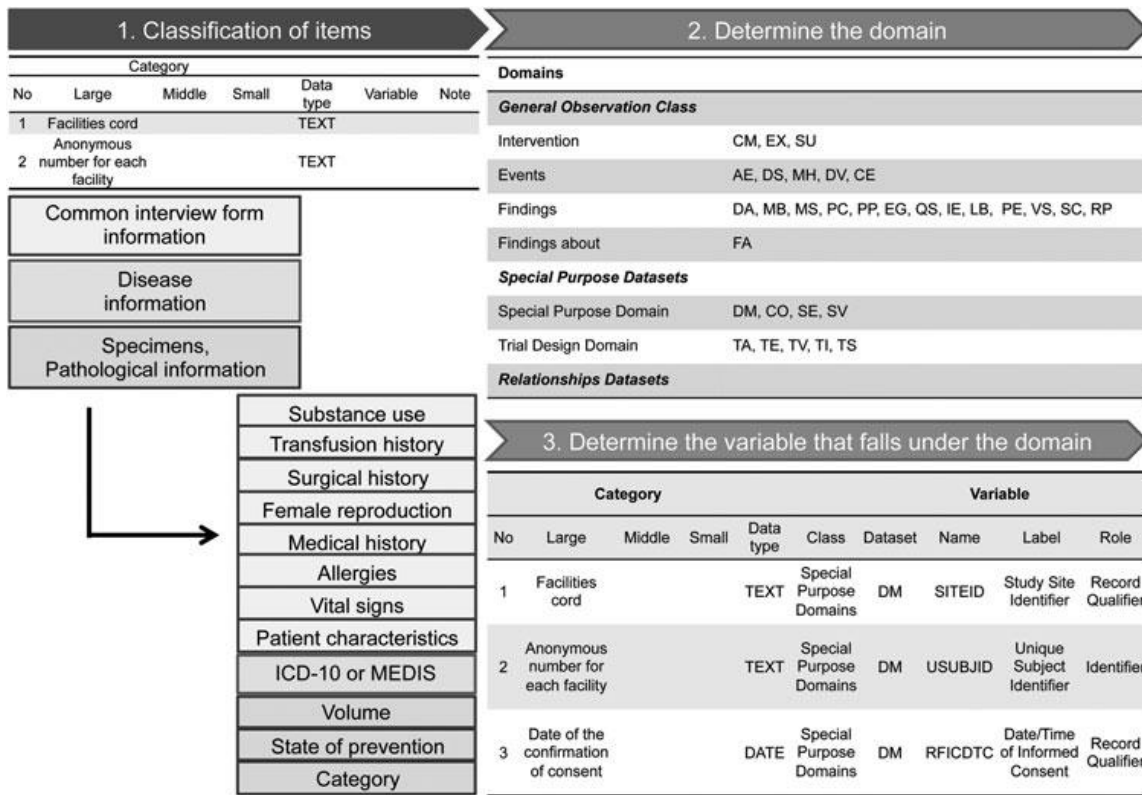
	2014	2015	2016
Biobank	← Step 1 →	← Step 2 →	
JCRAC (Clinical Data)	← Step 1 →		← Step 2 →

研究課題名 : 26指114 臨床データのCDISC標準化に関するフィージビリティスタディ
 主任研究者名 : 大津 洋
 分担研究者名 : -
 キーワード : CIDSC, バイオバンク, 臨床データ, データマネージメント, 標準

BioBankにおけるフィージビリティ研究(2014-2015)

(Step1,2)CDISC SDTM とバイオバンクのデータフォーマットのマッピングの実施 (Gap分析および解析可能性の検討)

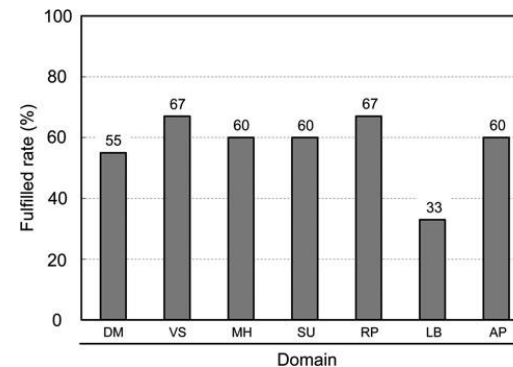
NCGN中央管理部門が持っている個票データフォーマットとのCISC SDTMへのマッピングマッピングモデルを策定し、その程度適合するのかを検討した。



結果 : 定義したマッピングでの適合容易性は以下ようになった。

Item name	Categorized domain of SDTM*		Easily classified		Problematic but classifiable	
	n		n	%	n	%
Common interview from information						
Facility/patient characteristics	9	DM Demographic	4	44.4	5	55.6
Vital signs	5	VS Vital signs	4	80.0	1	20.0
Medical history (patient)	27	MH Medical history	0	0.0	27	100.0
Surgical history	9	MH Medical history	0	0.0	9	100.0
Transfusion history	1	MH Medical history	0	0.0	1	100.0
Allergies	12	MH Medical history	2	16.7	10	83.3
Medical history (family)	30	AP Associated persons	0	0.0	30	100.0
Tobacco	8	SU,AP Substance use, Associated persons	3	37.5	5	62.5
Alcohol	7	SU Substance use	1	14.3	6	85.7
Female reproductive information	9	RP Reproductive system findings	4	44.4	5	55.6
Disease information						
Disease name	21	MH Medical history	0	0.0	21	100.0
Specimen, pathological sample information						
Specimen	51	LB Laboratory test results	18	35.3	33	64.7
Pathological sample	13	LB Laboratory test results	3	23.1	10	76.9
Total	202		39	19.3	163	80.7

*Please refer "Selected domains."
 SDTM, Study Data Tabulation Model.



また、全体としては、おおよそ全体の7割程度の項目は変換可能であるが、ドメイン別にみても、バラつきがあることが分かった(図)。同時に、治験では使用しないバイオバンク独自の検体に関する項目など、独自にSDTM用に変数を定義する必要があることが分かった。

考察

計画時に想定していたバイオバンクデータに関する国際的な標準化の目立った動きはなく、ましてやCDISCというキーワードが飛び交う状況ではないという判断のもと、臨床試験や治験を実施する上で、バイオバンク試料等を利用するという局面が出てくるのであれば、むしろそのような展開になるべきと考えるが、それを見越した地ならしの一歩となったと考える。

研究課題名 : 26指114 臨床データのCDISC標準化に関するフィージビリティスタディ
 主任研究者名 : 大津 洋
 分担研究者名 : -
 キーワード : CIDSC, バイオバンク, 臨床データ, データマネジメント, 標準

REDCap の導入とフィージビリティ研究

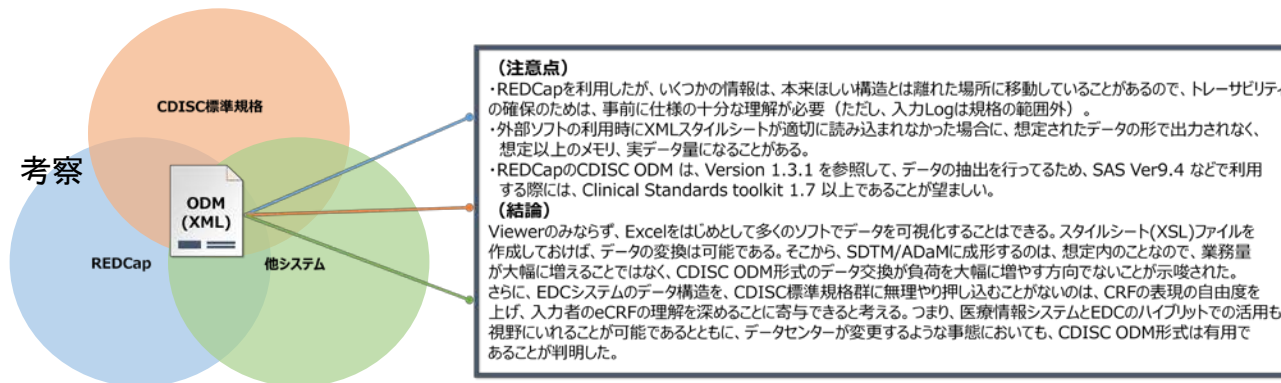
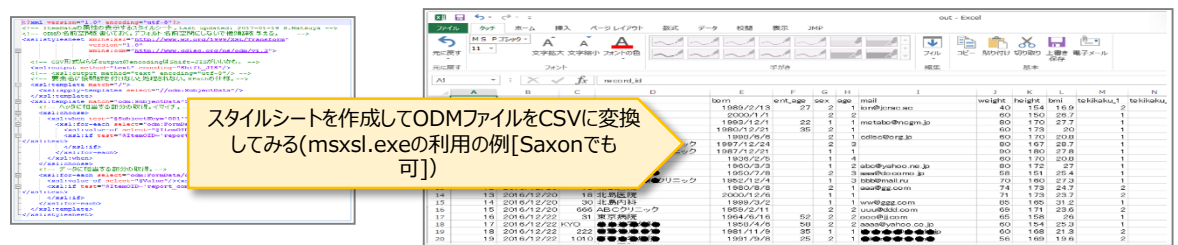
(Step1) REDCapの導入とデータの利活用について

REDCapのJCRACデータセンターへの導入とともに、CDISC形式と医療情報の厚労省標準規格SS-MIXについて、REDC介してデータの入出力が可能かどうかを実施。

Input example		Context
Type1	Using SS-MIX Data	SS-MIXデータは基本CSVになっているため、文字タイプとプロジェクトのフォーマットだけ注意すれば取込み可能。しかしHL7形式のフォーマットではすべて作り替える必要があった。
Type2	Using CDISC ODM	HISからODMフォーマットは出力できない。REDCapのODMフォーマットに合わせて作成を試みるが取込みはエラーとなる。REDCap Ver6.10.3では取込みできなかった。
Type3	Using CSV_File	問題なし。文字タイプとプロジェクトのフォーマットだけ注意すれば取込み可能だった。
Output Example		
TypeA	Using CDISC ODM/SDTM	HISからODMフォーマットは出力できない。REDCapのODMフォーマットに合わせて作成を試みるが取込みはエラーとなる。
TypeB	Using (SAS/SPSS)CSV etc..	問題なし

(Step2) EDCap上で、CDISCの一連のデータ交換・分析が可能かを調査・研究を行う。

- ・REDCapデータ (CSVファイル) とODMファイルの検証&ODMファイルの再利用
 REDCapからCSVとODMを抽出し、抽出されているデータに関して差異が無いことを検証した。プロジェクトから出力したODMファイルは、新規プロジェクト作成時にCRFとデータはすべてREDCapにアップロードできることが検証できた。(一部割付設定、サーベイ設定は不可)
 これにより、REDCapではODMファイルの扱いをone-wayでなくtwo-wayで行うことができる。
- ・REDCapから抽出されたODMファイルが、CDISC標準規格に基づいているか検証
 REDCapからODMファイルを抽出した場合、CDISC標準規格に定義されている、AdminData定義 (管理情報: 研究者やデータマネージャー、施設名、署名等) が出力されない。REDCapの場合はStudyのタグ内に記載されていた。他のタグについては、ODMに定義されている内容で出力できていることが検証できた。
- ・ODMファイルの利活用
 MS ExcelにてXMLファイルデータをデータ形式で確認することは可能であるが、不要な情報も同時に表示されている。スタイルシートを作成することによって、CSV形式やその他の形式に変換も可能なため、これまでのデータ変換のプログラムがあれば、それを有効活用することも可能。SAS proc xmlも利用可能だが、サポートしているODM Versionに注意が必要。



本研究の実施中に、医療機器の国際標準化 (行政) 団体であるIMDRFに参加し、レジストリデータベースの利用WGの参画とドキュメントの作成に寄与した。

研究発表及び特許取得報告について

課題番号：26指114

研究課題名：臨床データのCDISC標準化に関するフィージビリティスタディ

主任研究者名：大津 洋

論文発表

論文タイトル	著者	掲載誌	掲載号	年
Clinical Data Interchange Standards Consortium Standardization of Biobank Data: A Feasibility Study	Izumi Sato, Yohei Kawasaki, Yasuhiro Tanaka et al.	BIOPRESERVATION AND BIOBANKING	Vol. 14 No. 1	2016
研究者主導研究におけるCDISC標準規格群への対応	大津 洋	薬理と治療	45巻5号	2017

学会発表

タイトル	発表者	学会名	場所	年月
臨床研究におけるICT標準化の取り組みと現状・将来像について	大津 洋 田中 康博 岩本 智子 豊田 あこ 北島 浩二	第7回日本臨床試験学会	名古屋	2016年2月
研究者主導研究におけるCDISCへの対応	大津 洋	第16回 CRCと臨床試験のあり方を考える会議 2016 in 大宮	大宮	2016年9月
Feasibility study in REDCap using standard data structures	Hiroshi Ohtsu, Ako Toyoda, Koji Kitajima	REDCapCon	NC, USA	2016年9月
臨床研究におけるCDISC標準規格群利用に関する研究	大津 洋 北島 浩二 岩本 智子 豊田 あこ	第8回日本臨床試験学会	大阪	2017年1月
Real world evidence -戦略的レジストリネットワーク	大津 洋	療レギュラトリーサイエンス研究所 (IMeRS) 公開シンポジウム	東京	2017年3月

研究発表及び特許取得報告について

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
該当なし				

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
該当なし				

※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。