

prepopulating of CRF by the data in EMR has not yet been in practical use in actual clinical research. We developed a system, which transcribes some data required for a CRF from EMR automatically and sends completed CRF via network. We call this system as “eCRF reporter”. In this paper, we describe the outline of this system.

1. Methods

1.1. Data flow of the system

The purpose of eCRF Reporter is to generate an electronic CRF file in EMR system and send it to a clinical data management system (CDMS) in a datacenter. The interface module of eCRF reporter can retrieve the data in EMR database including patient biography data, laboratory test data, prescription data and data entered by templates in progress notes. The eCRF Reporter also enables users to enter data directly to eCRF. The eCRF Reporter generates a CDISC ODM file and a PDF document which is a readable form of Clinical data in ODM. After storing eCRF in EMR, it is transferred via VPN to the CDMS which can receive the eCRF files. When a data manager accepts the eCRF, the data in the ODM is stored into a database. The data flow concerning the eCRF Reporter is depicted in Figure 1.

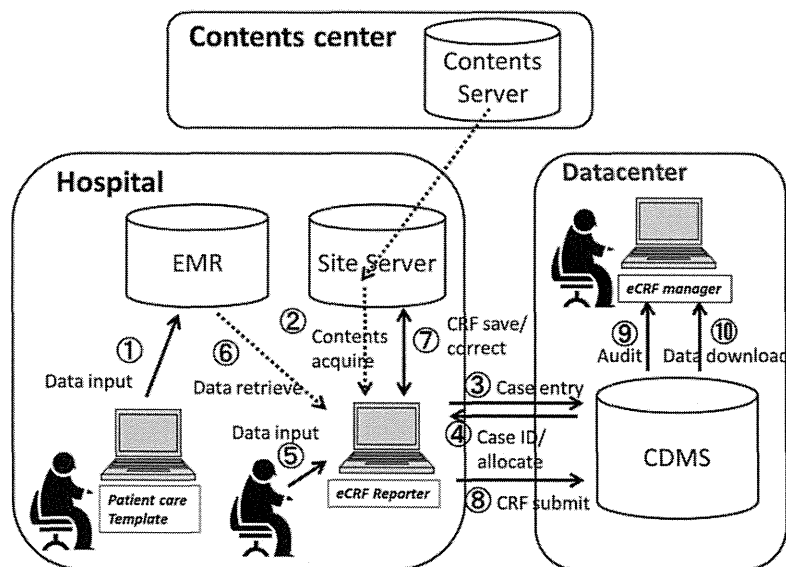


Figure 1. Data flow of the system

1) When doctors intend to write progress note, they use patient care templates to generate structured data²⁾. It outputs both narrative text for progress note field by using natural language generation technique and data XML for data processing. The data XML is essentially a set of key values. Each key item whose value is supposed to be retrieved from eCRF Reporter is put a reference code.

2) eCRF Reporter also uses the template module as user interface. Its task is to generate ODM file. Master data and program are completely separated in eCRF Reporter, so any

CRF can be made by letting eCRF Reporter program read master data. Adding to template master XML file, project configuration file and ODM file are also necessary for eCRF Reporter. These contents are delivered from a contents center to eCRF Reporter systems located in hospitals. By this mechanism, the hospitals which participate in a clinical research can generate the same CRF for each project.

3) When a new study case is found, a reporter creates study case with necessary information. At this time, the eCRF Reporter sends case information but personally-identifying information.

4) When the CDMS in the datacenter receives study case information, it sends back study case ID number. If the study is clinical trial and needs random allocation, the CDMS allocates the case into the prepared groups. The eCRF Reporter receives study case ID number and stores it in a local database.

5) A doctor or CRC inputs data by using CRF template for each study case.

6) The CRF template can retrieve data in the EMR database. The CRF Reporter program passes keywords to the interface module which is made by EMR vendor. Then the interface module retrieves patient's biography data, laboratory test data, prescription data, or data entered by the template in progress note from the corresponding database in EMR according to the keywords and returns retrieved data to the template engine.

7) Incomplete CRF can be saved in the local server and can be added or corrected afterward. After completing the CRF, the eCRF Reporter generates ODM file using data binding technique. It also generates a readable form of Clinical data in ODM as a PDF document. The ODM and the PDF are stored in EMR just before submission.

8) eCRF Reporter submits a completed eCRF (ODM and PDF) to CDMS in a datacenter via network.

9) eCRF Manager is client software of CDMS. Data manager can check the received eCRF using eCRF Manager. If the CRF contains some errors, the data manager refuses it and asks the CRC or the doctor in charge to correct it. If the data manager accepts it, the ODM is parsed and the containing data are saved into the CDMS database.

10) The data manager can download the data stored in the database of the CDMS to make an analysis of them.

1.2. System configuration and network

There are two types of installation configuration of the eCRF Reporter. One is server client type. The site server in a hospital keeps the contents of eCRF Reporter. When a user selects a study from a PC in the EMR system, the contents files corresponding to the study are delivered to the PC. The site server also has a database that stores the data XML, the ODM file and a PDF document generated by the eCRF Reporter. It also has a function of a proxy server to relay a completed eCRF sent from the PC to the CDMS. This type system needs always-connected network. We adopted On-Demand VPN provided by NTT data for it.

Another type is software that runs on a single PC. It has local database in the PC disk or USB memory. This type can be used anywhere without set up. Because the eCRF Reporter in this configuration is separated from EMR system, it cannot retrieve data from EMR database. Users connect with the network when they request the

contents files corresponding to the study and when they send entry information to the CDMS to get case ID. When users enter data to fulfill the eCRF, it can be used in offline mode. After finishing data entry, users again connect with the network to send eCRF files to the CDMS. For this type system, we adopted SSL-VPN provided by Cisco.

2. Results

The datacenter was set up in Osaka University Hospital, which deals with clinical studies conducted by Osaka University Medical School and its related facilities. By using this proposed system, 5 studies already started and 4 studies is now preparing. The outline of the 5 started studies is as follows: 1) Observation study with follow-up of acute coronary insufficiency. Its CRF includes 4 event forms. 2) Observation study with follow-up of mammary gland cancer. Its CRF includes 12 event forms, some of which are repeatable. 3) Clinical trial concerning preventive effect of clipping on breeding after colon polypectomy. This study uses stratified random allocation function of the CDMS. Its CRF includes 5 event forms. 4) Ovarian cancer registry. Its CRF includes 1 event. 5) Melanoma registry. Its CRF includes 1 event.

Just before starting this service, the system has been evaluated in detail and is regarded with high esteem by all of CRCs and doctors in charge

3. Discussion

When paper based medical record is used in daily practices, EDC system is the best way to promote efficiency in clinical research activities. However, in the clinic where EMR is used, the current EDC system has several shortages. Firstly, double data entry is required for daily care and clinical research. It is not only cumbersome but also mistakable. Secondary, the data entered for a clinical research into an EDC does not exist in the institutional database. Thus if a patient is enrolled into another research which requires the same data entered for the first study, even though, they have to be reentered into another EDC. The purpose of our project is realizing single data entry for care and clinical researches through EMR-CDMS data integration.

Our near-term target is multicenter academic clinical researches. Because contrary to sponsor initiated clinical trial, the budget of an academic clinical research is limited, they sometimes have to be conducted without employing CRC. In this case, double data entry for care and clinical research becomes a significant barrier for its execution. Our method can solve this problem. Academic research often includes observational study whose collecting items in CRF are more diverse than those of clinical trials; however, the items collected for observational study are almost same with those recorded in progress note in daily practice. Thus our method is more suitable for observational study than clinical trial.

The concept of our system has been already discussed well. CDISC Electronic Source Data Interchange (eSDI) Group proposed 5 scenarios for electronic source data (eSource) interchange³⁾. Our method corresponds to the 4th scenario; “eSource extraction and investigator verification (using electronic health records)”.

The implementation scheme of our project is similar with that of Retrieve Form for Data-capture (RFD) integration profile⁴⁾. “Form Manager” is compatible with the contents center in our system, which deliver ODM, template master and project configuration file to each site server. “Form Filler” corresponds to the template program with the interface module that creates eCRF. “Form Archiver” is one of the functions of EMR. “Form Receiver” is compatible with CDMS in the datacenter. However, the adopted concrete technique in our system especially for “Form Filler” is different from those in RFD.

The method we took is quite similar with the RE-USE project to the EHR4CR platform⁵⁾. In the RE-USE project, patient care centric templates were made independently to clinical researches. Thus a standardized terminology is necessary for re-using data. They used SNOMED for it, however, only 13.4% of the data elements in CRF were found in EMR templates. In our project, patient care centric templates are made simultaneously with templates for a clinical research. Thus we can set the common items with the same options in both templates and put reference codes beforehand. Thus standardized terminology is not indispensable for our system. In care templates, some items can be added if they are needed for daily practice. We assume that if a patient is enrolled to a clinical research, the doctor in charge uses special templates during research execution. Even though, the doctor can input any information not related to the research into the progress note.

Although our method can prepopulate CRFs without fault, the current system has a shortage especially when it will use in sponsored initiated clinical trials. It needs a lot of time to set up patient care centric templates in each hospital. Thus it may not be practical when many hospitals and clinics simultaneously participate in a clinical research. We are now planning to develop next version system in which master of patient care centric templates can also be delivered from a contents center. The doctors participating clinical research use the same template for patient care during clinical research execution period. We think this method will become more practical for re-use of data entered for daily practice for a clinical research.

References

- [1] Clinical Data Interchange Standards Consortium (CDISC) (Internet)]. 15907 Two Rivers Cove, Austin, Texas, USA; ©2008. Standard available from: <http://www.cdisc.org/models/odm/v1.3/index.html> (cited Apr 20, 2009).
- [2] Matsumura Y, Kuwata S, Yamamoto Y, Izumi K, Okada Y, Hazumi M, Yoshimoto S, Mineno T, Nagahama M, Fujii A, Takeda H. Template-based data entry for general description in medical records and data transfer to data warehouse for analysis. *Stud Health Technol Inform* 2007; 129 (Pt 1): 412-6
- [3] CDISC Electronic Source Data Interchange (eSDI) Group. Leveraging the CDISC standards to facilitate the use of electronic source data within clinical trials version, November 20, 2008. < <http://www.cdisc.org/stuff/contentmgr/files/0/2f6eca8f0df7caac5bbd4fadfd76d575/miscdocs/esdi.pdf> >
- [4] IT Technical Committee. Retrieve Form for IHE Data Capture (RDF) Trial Implementation IHE IT Infrastructure Technical Framework Supplement, August 10, 2010, http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_RFD_Rev2-1_TI_2010-08-10.pdf
- [5] El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic PY, Jaulent MC, Daniel C. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. *J Biomed Inform.* 2011 Dec 44; Suppl 1: S94-102

レセプトデータを活用した患者病名の推定

武田 理宏 三原 直樹 真鍋 史朗 松村 泰志

大阪大学大学院医学系研究科 医療情報学

Estimation of patient disease name using health insurance claims data in Japan

Takeda Toshihiro Mihara Naoki Manabe Shiro Matsumura Yasushi

Department of Medical Informatics, Osaka University Graduate School of Medicine

In constructing the clinical trial network, grasping the number of patients with each disease treated in the hospital is effective to assign the number of patients to be handled in the trial for the participated hospital. For this purpose, the data that is available in every hospital had to be used. In this study, we tried to develop the system to estimate the number of patients with each disease using receipt data. Because reliability of disease name registered to the receipt is very low, we estimate the disease name by using therapeutic medicine from receipt data. If it is difficult to estimate the disease name in detail (e.g. gastric ulcer), we should estimate rough disease name (e.g. gastrointestinal disease). For this, we allocated large-middle-small classification code (disease code) to each disease. Next, we create correspondence table between disease code and therapeutic medicine. Ten point was allocated to each disease code corresponded to one therapeutic medicine. Four point was added if disease code was registered to the receipt. Disease score was aggregated for each patient and we adopted the disease code of max disease score with 6 or more points. Next we exclude the adopted disease code and its therapeutic medicine, and same process was repeated to estimate the disease. First, we test 2 patients to estimate their disease name. We estimate 12 disease name from 18 therapeutic medicine, and 9 disease name from 12 therapeutic medicine. Only 1 disease name was considered not to be appropriate and 3 disease name could not be estimated. Next, we estimate disease name from receipt data from December 2013 at our hospital. 16,060 disease name was estimated from 19,921 patients receipt data (2,031 inpatient and 17,890 outpatient). 326 disease could be estimate to small classification (total of 10,866 disease), 40 disease to middle classification (total of 1,423 disease) and 16 disease to large classification (total of 3,771 disease). In conclusion, we could estimate disease name from disease data and therapeutic medicine data from receipt with a certain degree of reliability.

Keywords: disease name, health insurance claims, clinical trial

1. はじめに

臨床試験を行う場合、被験者を確保することが成否の鍵を握る。日本の場合、ある疾患を特定の医療機関に集中させるようなコントロールはされていない。このため、個々の医療機関では、十分な数の被験者が得られないことが多い。そこで、治験ネットワークを構築し、複数病院で仮想的に大規模病院を構成し、それぞれの臨床試験に対して十分な数の被験者を確保する体制作りが求められている。

治験ネットワークを構築する上で、各医療機関が診療する各疾患の患者数が分かると、治験の各医療機関への担当配分を決める際に有効である。近年、多くの病院で電子カルテが普及し、そのデータをDWHに保管することで、様々なデータ検索が可能となっている。同一施設内であれば、各疾患の患者数を求めることは難しくない。近年、医療データの標準化が進んでいるものの、多くの病院ではハウスコードを使用するなど、統一したデータ構造となっていない。治験は規模の異なる複数の病院が参加し、場合によっては電子カルテが導入されていない病院が参加することもある。臨床試験のニーズを満たすためには、複数の医療機関で統一的な方法で情報検索できなければ意味をなさない。現状では、レセプトオンライン化が制度化され、レセプトデータ、DPCデータはコード、形式が標準化されており、かつ、各病院から容易にデータを出すことができる。レセプトデータのみで各疾患の患者数を推定出

来ることが出来れば、このニーズを満たすことが可能となる。

レセプトデータは、病名が含まれているが信頼性に乏しいことが問題である。また、検査結果等の診断の根拠となるデータが含まれていない。しかし、投薬内容や処置、手術等の治療内容については高い精度で含まれている。DPCデータは病名を含め信頼性が高いが、外来患者に適用できない問題がある。こうした特性、限界を理解することで、患者数の概数を求めるシステムを構築することができる可能性がある。

本研究では、こうした考えのシステムの実用性を調べるために、大阪大学医学部附属病院のレセプトデータ、DPCデータを取得し、解析しやすいデータベースに展開して、患者数の概数を求めるシステムを開発した。本システムが期待通りに機能することが確認できた場合は、各病院に展開し、統計データをネットワーク事務局に報告するシステムに拡張することが可能である。

2. 方法

2.1 ICD-10コードの疾患コードへの変換

病名は可能であれば詳細な病名まで推定する方が好ましいが、詳細な病名まで推定出来ない場合は、粗い分類であっても病名が推定出来ることが望ましい。ICD-10コードは標準化された疾患コードであるが、疾患の上位下位概念が明確でない。そこで大分類、中分類、小分類をもつ疾患コードを作成し、ICD-10コード(3ケタ)を「疾患コード」に変換した。例えばI20:狭心

症を、大分類:09循環器系の疾患、中分類:04虚血性心疾患、小分類:01狭心症、I21:急性心筋梗塞を、大分類:09循環器系の疾患、中分類:04虚血性心疾患、小分類:02急性心筋梗塞に変換し、可能であれば小分類:狭心症、急性心筋梗塞まで推定を行うが、難しければ中分類:虚血性心疾患、中分類も難しければ大分類:循環器系の疾患の推定を行うこととした。

2.2 「医薬品と対応病名データ」による候補病名の抽出

レセプトに記載される薬剤については、一般財団法人日本医薬情報センター(JAPIC)が用意する「医薬品と対応病名データ」を活用した。「医薬品と対応病名データ」は、医療用医薬品添付文書の効能効果とこれに対応するICD10 対応標準病名マスター(MEDIS 標準病名マスター)を関連付けしたもので、使用した薬剤とレセプトに記載される病名の妥当性を評価するために使用されている。「医薬品と対応病名データ」には評価結果が、1:標準病名として妥当と判断したもの、3:標準病名として妥当性に判断を要するもの、5:標準病名として妥当性を欠くもの、の三段階で表記されている(表1)。本研究では「医薬品と対応病名データ」で変換されるICD-10コードの3ケタのみを使用し、「疾患コード」に変換した。

表1 「医薬品と対応病名データ」の実例

薬剤レセコード	告示名称	ICD10	疾患名	評価結果
610444042	ガスターD錠	K21	胃食道逆流症	1
610444042	ガスターD錠	K25	胃潰瘍	1
610444042	ガスターD錠	K26	十二指腸潰瘍	1
610444042	ガスターD錠	K27	部位不明の消化性潰瘍	1
610444042	ガスターD錠	K28	胃空腸潰瘍	1
610444042	ガスターD錠	K29	胃炎および十二指腸炎	1
610444042	ガスターD錠	K31	ディスペプシア(症)	1
610444042	ガスターD錠	K63	腸のその他の疾患	3

2.3 疾患点数による病名の推定

レセプトに記載された薬剤に対し、それぞれ10点の点数を与えた。JAPICの「医薬品と対応病名データ」によって与えられる候補疾患(小分類)に対し、「(評価結果:1):(評価結果:3)=2:1」で重みづけをしたうえで、薬剤ごとに与えられた「疾患点数」10点を配分し、「疾患点数(小分類)」を与えた(表2)。

大分類、中分類が共通する候補疾患については、「疾患点数(小分類)」を加算し、「疾患点数(中分類)」を求めた。また大分類が共通する候補疾患については、「疾患点数(中分類)」を加算し、「疾患点数(大分類)」を求めた。次にある患者に対し複数の薬剤が使用される場合、同一疾患分類ごとに疾患点数を加算し、使用した全薬剤に対する疾患点数(小分類)、疾患点数(中分類)、疾患点数(大分類)を求めた。

レセプトに記載される病名は、いわゆる保険病名、終了を忘れた病名などがあり信頼性は高くないが、実際に患者が罹患している病名が登録されていることも少なくない。このため、病名の推定に対しては、レセプトに記載された病名も活用するべきである。そこで、レセプトに記載される病名を「疾患コード」に変換し、それぞれの疾患コードに対し4点の点数を与えた。使用した薬剤に対する疾患点数と、レセプトに記載された病名に対する疾患点数を加算し、患者に対する疾患点数(小

分類)、疾患点数(中分類)、疾患点数(大分類)を求めた(図1)。

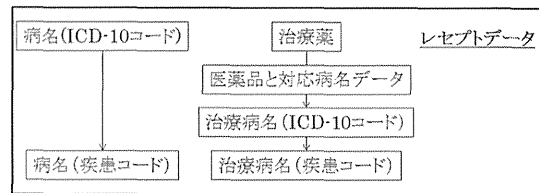


図1 疾患の推定方法

表2 ある薬剤に対する候補疾患と疾患点数配分例

大分類	中分類	小分類	評価結果	重みづけ	疾患点数	小分類	中分類	大分類
1	1	1	1	2	2.5			
1	1	2	1	2	2.5			
1	1	...					5	
1	2	2	1	2	2.5			
1	2	...					2.5	
1	...	...						7.5
2	1	2	3	1	1.25			
2	1	3	3	1	1.25			
2	1	...					2.5	
2	...	...						2.5
						10	10	10

このように使用した薬剤とレセプトに記載された病名に対し与えられた疾患点数が高いことは、実際に患者がその疾患に罹患している可能性が高いことを意味する。逆に疾患点数が低いことは薬剤情報と病名情報だけでは疾患を推定できないことを意味する。そこで疾患採択閾値=5とし、疾患点数(小分類)で採択閾値以上で最も高い疾患点数となった疾患を採択疾患とした。採択疾患を候補疾患コードに含む薬剤と病名を除外し、再度、疾患点数の計算を行い、採択閾値以上で疾患点数(小分類)が最も高い疾患を採択疾患とする作業を繰り返した。全ての疾患点数(小分類)が採

1-F-2-3 一般口演/1-F-2:一般口演7

採択閾値未満となった場合、疾患点数(中分類)が採択閾値=5以上で最も高い疾患を採択疾患とし、全ての疾患点数(中分類)が採択閾値未満となった場合は、疾患点数(大分類)で同様の検討を行った。全ての薬剤に対し、疾患コードが割り振られるか、疾患点数(大分類)が採択閾値より小さい値となるまで、処理を繰り返した。

3. 結果

3.1 病名推定の具体例

大阪大学医学部附属病院の実患者2名に対して、病名の推定を行った。

3.1.1 症例1

症例1は、①「拡張型心筋症」に対して、β遮断薬投与中。②「心房細動」、「心原性脳梗塞」にてワーファリン投与中。③「高度肥満」、「糖尿病」があり、インシュリン治療中。④「脂質異常症」に対し、スタチン投与中。⑤下血から貧血が出現し、大腸内視鏡にて「大腸がん」と診断。⑥消化器外科にて大腸切除術を施行した。⑥手術に際して、ワーファリンからヘパリン置換を行った患者である。

投与薬剤は、外来にてコニール錠、アテレック錠、ディオバン錠、アーチスト錠(2.5mg)、ジゴキシン錠、バイアスピリン錠、ワルファリンカリウム錠、リピトール錠、ラシックス錠、タケプロンOD錠、ヒューマリンR注を使用し、入院にてレンドルミン錠、カロナール錠、プルゼニド錠、生理食塩液、ビーフリード輸液、ヘパリンナトリウム注、フェンタニル注射液を使用した。

レセプトには25病名が登録され、うち9病名が治療中の疾患として妥当と考えられた。全投与薬剤18剤に対して12病名が採択疾患として割り振られた。うち9病名は小分類までの推定、3病名は中分類までの推定であった。6病名はレセプトに病名として登録されていなかったが、主治医により疾患名として妥当と判断された。レセプト病名として妥当と医師が判断したものうち、採択疾患として採択されなかったのは、(03,01,01)鉄欠乏性貧血、(04,08,09)脂質異常症、(09,06,19)心房細動であった。鉄欠乏性貧血については前月は鉄剤を投与されており、採択疾患となったと考えられるが、対象月は鉄剤の投与がなかったため採択疾患とならなかった。脂質異常症については(04,02,05)糖尿病として、心房細動については(09,07,04)脳梗塞として疾患採択されたため、採択疾患から漏れた。

3.1.2 症例2

症例2は、①「慢性関節リウマチ」、「混合性結合組織病(MCTD)」に対してステロイド、メトトレキサート、アバセプト投与中。②「肺高血圧症」に対し、ボセンタン、プロサイリン、サルボグレラート投与中。③「鉄欠乏性貧血」に対し、鉄剤投与中。④「脂質異常症」に対し、スタチン投与中の患者である。

投与薬剤はアンブラーグ錠、ドルナー錠、フェロミア錠、リウマトレックスカプセル、ラシックス錠、ミカルデイス錠、クレストール錠、トラクリア錠、アルダクトンA錠、メドロール錠、生理食塩水、オレンシア点滴静注用を使用した。

レセプトには19病名が登録され、うち7病名が治療中の疾患として妥当と考えられた。全投与薬剤12剤に対して9病名が採択疾患が割り振られた。うち8病名は

小分類まで、1薬剤は中分類までの推定であった。3病名はレセプトに病名として登録されていなかったが、生理食塩水に対する病名は妥当と判断されなかった。生理食塩水については、オレンシア点滴静注用の希釈液として使用したもので、そもそも疾患に対する投与でなかったためである。メドロールは非常に多くの候補疾患があったが、(13,02,02)慢性関節リウマチを正しく推測出来た。ラシックス錠、ミカルデイス錠、アルダクトン錠は(09,06,21)心不全に対して投与されていたが、(09,03,01)高血圧に対する投与と推定された。また、(13,02,02)慢性関節リウマチに点数を取られたため、(13,05,05)全身性硬化症、(13,05,06)その他全身性結合組織疾患が推定されるべき疾患から漏れた。

3.2 構築した患者数推定プログラムによる疾患の推定

構築した患者数推定プログラムに対し、2013年12月のレセプト電算、EFデータより評価データを作成し、データ内容の確認を実施した。処理対象データは患者数、入院2,031人 外来17,890人に対し、EFファイルが入院156,296レコード、外来169,500レコード、レセプト電算が入院7,976レコード、外来134,140レコードであった。

患者数19,921人から16,060件の疾患名が推測された。うち疾患コードの小分類まで推測出来た疾患名は326疾患、10,866件、中分類までの推測は40疾患、1,423件、大分類までの推測は16疾患、3,771件であった。小分類まで推定された疾患で件数が多かった10疾患を表3に示す。最も件数が多かったのは(04,02,05)詳細不明の糖尿病で558件であった。

表3 疾患コードの小分類で推定された疾患とその件数(上位10疾患)

	大分類	中分類	小分類	疾患名	件数
1	4	2	5	詳細不明の糖尿病	558
2	4	8	9	リポたんぱく代謝障害 およびその他の脂血症	526
3	6	6	7	睡眠障害	470
4	11	6	5	その他の腸の機能障害	370
5	13	12	2	骨粗しょう症、 病的骨折を伴わないもの	327
6	7	6	1	緑内障	315
7	13	2	2	その他の慢性関節リウマチ	307
8	9	3	1	本態性高血圧	270
9	10	4	1	血管運動性鼻炎	268
10	4	2	2	インスリン非依存性糖尿病	217

疾患コードの小分類まで推定された疾患を含め、中分類で最も件数が多かった疾患は(04,02,**)糖尿病で948件、ついで(04,08,**)代謝障害が816件であり、代謝性疾患が上位を占めた(表4)

表4 疾患コードの中分類で推定された疾患とその件数(上位10疾患)

	大分類	中分類	小分類	疾患名	件数
1	4	2	**	糖尿病	948
2	4	8	**	代謝障害	816
3	11	2	**	インフルエンザおよび肺炎	712
4	6	6	**	挿聞性および発作性障害	688
5	13	2	**	炎症性多発性関節障害	658
6	13	12	**	骨障害および軟骨障害	435
7	11	6	**	腸のその他の疾患	424
8	9	3	**	高血圧性疾患	350
9	12	3	**	皮膚炎および湿疹	346
10	7	6	**	緑内障	315

最後に小分類、中分類まで推定された疾患を含め大分類での疾患推定件数を表5に示す。最も多かったのは(04,**,**)内分泌、栄養および代謝性疾患で2,087件あり、ついで(11,**,**)消化器系の疾患、が1,952件、(09,**,**)循環器系の疾患が1801件と続いた。実際に疾患出現頻度の多い疾患が上位を占めている印象を受ける。一方、(17,**,**)先天奇形,変形および染色体異常、(20,**,**)傷病および死亡の外因の疾患推定件数は0件であった。

	大分類	中分類	小分類	疾患名	件数
1	4	**	**	内分泌、栄養および代謝疾患	2087
2	11	**	**	消化器系の疾患	1952
3	9	**	**	循環器系の疾患	1801
4	6	**	**	神経系の疾患	1470
5	13	**	**	筋骨格系および結合組織の疾患	1461
6	5	**	**	精神および行動の障害	1352
7	1	**	**	感染症および寄生虫症	962
8	7	**	**	眼および付属器の疾患	909
9	2	**	**	新生物	762
10	12	**	**	皮膚および皮下組織の疾患	761

図5 疾患コードの大分類で推定された疾患とその件数(上位10疾患)

4. 考察

治験ネットワークを構築する上で、各医療機関が診療する各疾患の患者数を把握することは重要である。このためには、規模の違う複数の病院のデータから疾患数を推定する必要がある。電子カルテデータの二次利用はその候補となり得る。しかし、多くの病院で標準コードとハウスコードのマッピングが進んでいるだけでなく、現時点でその活用は困難である。レセプトオンライン化が推進された結果、レセプトデータは電子化されただけでなく、記載コードもレセ電コードで統一されている。

レセプトデータは、ここに記載されている病名に信憑性が低いことが問題とされている。既に治癒した病名が終了されていなかったり、疑いであるべき病名に疑

いの表記が抜けていたりする。そこで、本研究では、レセプトに記載されている治療内容に着目し、治療内容から、治療対象病名を推定する方法を検討した。

病名を推定する際に、病名概念の粒度を考慮することが重要である。ある患者の病名について、医師Aは細かな粒度の病名を付け、医師Bはやや粗い粒度の病名を付けたとする。例えば、医師Aは十二指腸潰瘍と診断し、医師Bは同じ患者を消化性潰瘍と診断するなどである。二人が付けた病名は異なっているが、本質的には一致している。これを解決するためには、個々の病名登録に対し上位概念の病名も同時に登録されていると捉える必要がある。この例では、十二指腸潰瘍の病名は、消化性潰瘍でもあり、胃疾患でもある。このようにすれば医師Aと医師Bは基本的には同じ病名を想定していたことが判定できる。

本法では、病名の階層概念をデータ化し、各治療薬について、同一レイヤーの治療対象疾患数で按分する方法をとっている。例えば、抗生剤はあらゆる感染性の疾患が治療対象疾患となるので、このレイヤーでの個々の疾患に対する配点は低くなる。しかし、全てが感染症であるので、感染性疾患の大分類粒度のレイヤーでは、感染性疾患が高得点となる。

本法は、医師が付けた病名に対して、治療がされている場合は、その病名を治療対象病名として拾い出すことを方針としている。そこで、医師が付けた病名に対しても配点し、病名だけでは基準値を超えず、その疾患に対する治療薬がある場合に、点数が加算され、その最大値をとる病名が最終的に拾い出す仕組みとなっている。本研究では患者数19,921人に対し、16,060件の疾患名の推測に留まっている。これは、治療薬が投与されていない、他院で治療薬が投与されている患者については、病名が採用されないことによる。今後、手術、処置、検査データを組み合わせることで、推定出来る病名の精度や件数を増やすことができる可能性がある。

本研究では、病院で診療する患者数を把握することを目的としている。複数の病院を同じ方法で処理することにより、複数病院を比較できることが重要であると考えている。ある臨床試験で、ある疾患を持つ患者が被験者の条件であったとする。本法で得たデータから、その被験者が何人いるかを推定することはできないが、例えば、治験ネットワークの複数病院で30例を受け持つ必要があった場合に、それぞれの病院の受け持ち数を、本法で得られた各病院のその病名を持つ患者数で按分して配分することには、それなりの合理性がある。

5. 結論

レセプトデータの投薬内容と病名から治療対象病名を拾い出し、リストアップするシステムを開発した。サンプル症例からある程度の精度で治療対象病名が拾い出せていると考えられた。

病院情報システムを治験データとして活用することへの展望と課題

青柳 吉博^{*1} 千葉 吉輝^{*2} 岡田 昌史^{*3} 赤堀 澄子^{*4} 溝渕 真名武^{*5} 横井 英人^{*4}

^{*1}国立がん研究センター ^{*2}東京大学病院 UMINセンター

^{*3}CDISC Japan Users Group CDASHチーム ^{*4}香川大学医学部附属病院医療情報部

^{*5}富士通株式会社ヘルスケア・文教システム事業本部ライフイノベーション事業部

The prospects and problem to utilizing a hospital information system as investigational data

Aoyagi Yoshihiro^{*1} Chiba Yoshiteru^{*2} Okada Masafumi^{*3} Akahori Sumiko^{*4}

Mizobuchi Manabu^{*5} Yokoi Hideto^{*4}

^{*1}National Cancer Center Center for Research Administration and Support

^{*2}University of Tokyo UMIN Center

^{*3}CDISC Japan Users Group CDASH Team

^{*4}Department of Medical Informations Kagawa University Hospital

^{*5}FUJITSU LIMITED, Healthcare/Educational Systems Unit, Life Innovation Div

The development of hospital information systems in Japan have spread nationwide, and the overall induction rates are 18.7% and 56.9% for 400 beds or more. It is becoming increasingly natural for clinical records and information to be saved electronically and the profit utilization of this saved electronic data is currently being examined. The advantages of usage of electronic database are an increase in business efficiency, globalization, and expansion for secondary use. Furthermore, shared clinical record information is possible to optimize values of the regional medical cooperation system. However, in the case of clinical trials to ensure the reliability of information, the cooperation of electronic intelligence needs to be performed. The data which is populated in the case report (CRF) is usually entered by hand into the EDC system, which a sponsor prepares. To address the issues mentioned above, it is necessary to consider how the trial data is managed by the hospital staff and the pharmaceutical companies, as well as what kind of approach is necessary for securing the reliability of the hospital information system for future use. Therefore, in this session, we will give an outline of the Clinical Data Interchange Standards Consortium (CDISC) standard, which is an information exchange agreement to use the hospital information system as trial data. In addition, we discuss the possibility of data usability to the pharmaceutical companies through the implementation of the Clinical Data Acquisition Standards Harmonization (CDASH), which is an agreement to define what data are collected in the CRF. Moreover, we discuss the computerized system validation (CSV) and the electronic record/e-signature (ER/ES) as areas for the medical information system to work on in order to secure the reliability of data when expanding to pharmaceutical companies. If seamless information sharing is enabled after having secured reliability, we will be able to relieve the burden on hospital staff and save costs for pharmaceutical companies. Consequently, this will contribute to the shortening of the process associated with drug development.

Keywords: CDISC, Computerized System Validation, Clinical Trial, Hospital Information System

1. はじめに

本邦における病院情報システム(HIS)の普及は全国的に広まっており、全体の導入率は18.7%、400床以上では56.9%となっている(2012年実績)。カルテ情報は電子的に保存されることが一般的となりつつあり、現在では保存された電子データの利活用に関する検討が多く行われている。

電子データの利点は情報共有による業務効率化やグローバル化、データの2次利用への展開が挙げられ、地域連携システムを用いたカルテ情報共有や臨床研究への利用などが盛んに行われている。一方で治験など信頼性保証を重要視する臨床試験の場合、電子情報の連携はほとんど行われておらず、現在でも症例報告書(CRF)へのデータ記載は治験依頼者が用意するEDCシステムに病院スタッフ(治験責任医師等や臨床研究コーディネーター(CRC)など)がデータを手入力する運用が一般的である。

上記に対して平成24年に発出された臨床研究・治

験活性化5カ年計画2012および同アクションプランではさらなるIT技術の活用が謳われており、医療機関における体制整備としてHISとEDCの連動やSS-MIXの利用などが挙げられている。(図1)

○短期的な目標

- ・治験審査委員会等の業務のIT化(審査資料の電子ファイル化等)
- ・EDC(Electronic Data Capturing)の利用の促進
- ・リモートSDV実施に向けた調査・研究

○中長期的な目標

- ・臨床研究中核病院等の臨床研究の中核的役割を担う医療機関においては、病院情報システムとEDCとの連動について取り組む。
- ・治験業務のIT化の基盤となるSS-MIX(Standardized Structured Medical Information Exchange)標準化ストレージやCDISC(Clinical Data Interchange Standards Consortium)標準等の導入を検討する。
- ・治験依頼者、医療機関は、費用対効果を勘案しながらクラウドコンピューティングの活用等について検討する。
- ・国は、一定のルールを設けた上で、産業界も含めて広く活用できる、大規模医療情報データベース(例えば厚生労働省の「医療情報データベース基盤整備事業」等)の在り方を検討する。

図1 臨床研究・治験活性化5カ年計画2012および同アクションプランにおけるIT化の目標

HISのデータがEDC等を介してシームレスに製薬企業に提供されれば、製薬企業はデータを取得・確認するための工数を削減することが可能となり、実施医療機関においてはデータの入力などに要する時間が削減されるなど、治験の効率化・質の向上に有用であると考えられる。

しかし、これらを実用的に運用した例は少なく、また適切なソリューションも見当たらないのが現状である。また、EDC連携を行うに当たりグローバルも含めた規制当局や治験依頼者に対応することを前提としたHISの信頼性確保に対して検討した事例も少ない。

そこで、本セッションではHISとEDCを連携させるために必要な規格であるCDISC標準について解説すると共にその実装事例について発表する。また、臨床試験における収集項目を定義する規格であるCDASHについて概説するとともに、CDASHを用いるための課題について考察する。

さらに、HISデータを治験に用いるために必要なコンピュータシステムバリデーション(CSV)に関する検討事例を紹介すると共に海外規制に対応するためアクションプランについても述べる。最後にHISベンダー側の意見としてEDCとの連携において必要なコンポーネントの実装事例を説明するとともに、今後のパッケージ開発におけるHISベンダーのマイルストーンについて述べる。

2. CDISC標準について

CDSICはClinical Data Interchange Standards Consortiumの頭文字略であり、その組織はベンダーやプラットフォームに依存せず、臨床研究データとメタデータの取得・交換・申請・アーカイブをサポートするための標準の開発とそれらのサポートをしている米国に本部を置く世界的でオープンな非営利の標準化団体である。

CDISC標準は、当初の目的であった治験の電子申請だけではなく、治験・アカデミックな研究も含めた広く一般の臨床・疫学研究において、

- 1)医療機関等における臨床研究データの収集
- 2)データセンターとの通信(EDC)
- 3)データセンターにおけるデータ管理
- 4)統計解析(メタアナリシスを含む)

等のあらゆる臨床・疫学研究の領域において、活用できる規格群に成長している。現在のCDISCのメンバー数は300以上の組織となり90以上の国と地域で活動している。最近SDTMを中心に疾患領域ごと(Therapeutic Areas)の標準の開発も積極的に進められている。

3. 大学病院医療情報ネットワーク研究センターにおけるCDISC ODMを利用した症例集積システム

大学病院医療情報ネットワーク研究センター(UMIN(University hospital Medical Information Network)センター)では研究者主導型臨床試験支援システムであるインターネット医学研究データセンター(INDICE(Internet Data and Information Center for Medical Research))を運用中であり、現在までに200以上の試験プロジェクトが実施され登録症例数は300万例以上となっている。同システムでは平成21年よりCDISC ODM形式での臨床データの電子的送付を開

始している。本システムは、実際の臨床試験において臨床データが各施設よりCDISC ODM形式で電子的に送付され、システムに自動登録することができるようになっている。データの交換方式にCDISC ODM形式を採用することで、システムが標準化され効率的な連携が出来るようになった。これにより病院情報システム中に存在する臨床データの治験データへの2次利用が促進されることを期待している。

CDISC ODMは治験データのデータ交換規約であり実際のシステムの通信に係わる部分は既定の範囲外となっている。医療機関とのデータ交換を考えると実際の通信を含む電子的データ送付方法を考えなければ現実的なデータ交換は不可能であり、その機能仕様が統一しない場合には結局相互接続に支障を来すことになる。そこでUMINセンターでは、臨床データのCDISC ODM形式での表現の実現と共に、INDICEにて実現した通信部分の仕様を公開しており、標準として普及を目指している。本発表ではCDISCの概要とUMINセンターでのCDISC ODM形式での臨床データ交換実装について説明する。

4. HISデータよりCRFを作成するためのCDASH

CDASHはCDISC標準の中でも臨床試験データの収集において推奨される収集項目を定義している。現在のCDASHはCRFデザインにおいて治療領域に依存しない共通なデータを収集する仕様となっている。収集されるデータの表現形式、メタデータがCDASHに準拠して標準化されれば、HISデータを治験データとして活用する際のデータマネジメント作業が省力化され、治験依頼者における試験の開発期間短縮・質向上に寄与できる。また、2013年6月に発表された「SS-MIX標準ストレージを活用した製造販売後の調査・臨床研究推進に関する提言」でも触れられているとおり、日本におけるCDISC標準の十分な浸透はSS-MIXからの標準的なデータ出力を大いに助けるものである。中でもCDASHは臨床試験データの収集項目を直接定めたものであるため、医療機関において臨床試験・治験などに携わる者はCDASHについて理解し導入することで大きな恩恵を享受できる可能性がある。また、治験実施時における症例報告書の作成において収集項目をCDASHとする動きは治験依頼者としてのトレンドとなりつつあり、先日発出された、「承認申請時の電子データ提出に関する基本的考え方について」の中でもCDASHの利用推奨が謳われていることから、HIS側においても対応できるだけの準備は必要と考える。

5. 治験における症例報告書とCDASHとの比較評価

実施医療機関から治験依頼者へは、症例報告書(CRF)を用いて実施結果が報告される。これらを電子的に扱うときの項目セットとして、国際的な臨床試験データ交換仕様のCDASHが準備されている事は前述の通りである。しかし、CDASHは現在も策定の途上であり、その粒度と範囲は項目により偏在しているため、必要な項目を十分にカバーしていない。その結果、治験ごとに独自の項目が設定され、データ項目の標準的な活用方法が確立していない事も問題となっている。そこで、香川大学病院では過去に行われた数例の治験プロトコル(膠芽腫、前立腺癌等)のCRFを対象と

し、実際に運用されたCRFの項目とCDASHで定義されている項目とを比較し、どの程度一致しているかを下記の方法にて調査した。

1)CRFを構成するデータを項目名ごとに抽出する。同じ項目名でも、データ収集タイミングが異なる場合は区別してカウントする。

2)CRFの項目に一致するCDASHのラベルを探す。名称が一致しない場合も、各ラベルの持つ意味を解釈し、同じ意味合いを持つラベルに分類する。

3)マッピング出来ない場合は、不一致項目に分類する。

その結果、平均してデータ項目の約6割がCDASHに対応していた。また、CDASHには画像検査・評価に関する項目が不足していることがわかった。今後は、CDASHでカバーできていない箇所を洗い出し、必要に応じてCDISCのルールに則った拡張を行い、CRFの標準入力様式の整備を試みる。さらに、CDASHの活用をベースとした治験依頼者やプロトコルに依存しないCRFの共通化を達成し、治験の電子化プロセスを促進し、治験の効率化やデータ品質の向上につながる仕組みの構築を目指す。

6. HISデータを治験に用いるために必要な信頼性の担保について

治験で収集する症例報告書の再現と評価のために実施医療機関ではその「原資料」を治験薬の承認申請後においても継続的に保管している。昨今、電子カルテの普及と共に、保管形態は紙から電子へと変化してきた。電子カルテシステム等の電子記録上にあるものを原本として取り扱うためには、ER/ES指針(本邦)および国際共同治験の場合などの場合にはFDAやEMAなど海外規制への対応が必要とされている。

最近の日本国内では、「医療情報システムの安全管理に関するガイドライン」を満たしていればER/ES指針への対応を満たすとみなし、電子カルテ上に保存された記録を治験の原資料として認める方向となっているが、安全管理ガイドラインへの準拠をすべての実施医療機関が保証しているわけではない。また、医薬品開発体制がグローバル化していることも踏まえて海外規制に沿った対応が必要と考えられる。そこで、国立がん研究センターでは、上記課題に対応するために、「電子カルテシステムの回顧的CSV」、「部門システムのCSV」および「原データ管理システムの構築(後述)」を行った。

これらの取り組みにより、部門システムを含めたHIS全体として治験の原資料の電子記録に対する信頼性が大きく向上し、システムの運用保守やリプレースを含めたライフサイクル全般における見読性や保存性が確保され、国際標準・規約への対応力が強化された。また、治験依頼者に対して信頼性が保証されたHISのデータが生成されることになり、EDCシステムとの連携やリモートSDVへの活用など、治験の電子化を大きく加速していくことが可能となった。

7. 治験原資料の電子化と連携活用に向けたHISベンダーの取り組み

本邦におけるHISは、電子カルテ部分はもちろんのこと画像や検査などを行う部門システムにも「原資料」に相当するデータが分散保管されている。部門システムはしばしばリプレースされることがあり、データ形式がリプレース前後で変わってしまうことや、データ移行の保証等、電子記録の見読性や保存性の維持に問題をきたす可能性があった。実際、富士通の電子カルテシステムは、データは部門システムで分散保管し、カルテ端末から各部門システムのビューワーを起動しているケースが多いため、システムのリプレース時にはデータの継続利用に細心の注意を払う必要がある。上記の問題を解決する手段として、富士通は国立がん研究センターと共に、電子カルテシステムと連動する部門システムデータを収集管理する「原データ管理システム」を新たに開発・導入した。

原データ管理システムは、各部門システムに保管されるオリジナルのデータからDICOMやPDFなどの標準的データフォーマットの形式で一元的にストレージに収集できるようにする仕組みと、共通の連携インターフェイスを実装した。収集するデータは画像等のデータそのものだけでなく、属性情報を合わせて収集し、GCP(医薬品の臨床試験の実施の基準)等で規定する「保証付き複写データ」(=Certified Copy)として収集し、治験の記録として原本性を担保できるようにした。属性情報としては、データの発生または更新日時・発生場所(オリジネーター)等の情報がおり、専用のビューワー機能で確認することができるよう検討した。

富士通では今回の取り組みをきっかけに、今後さらにHISデータの臨床研究・治験での活用を統合的・一体的に支援するソリューションを提供することで、HISの価値をさらに高め、医療イノベーションのためのインフラ充実に貢献していきたいと考えている。

