

Use of FHIR in Clinical Research: From Electronic Medical Records to Analysis

Jeff Abolafia, Rho, Chapel Hill, NC, USA
Priya Gopal, TESARO, Waltham, MA, USA
Sam Hume, CDISC, State College, PA, USA

ABSTRACT

In two previous papers, the PhUSE working group "Investigating the Use of FHIR in Clinical Research" demonstrated that data typically collected in diabetes studies can be extracted from medical records through FHIR (Fast Healthcare Interoperability Resources) and we can automate the process to populate eCRFs (electronic Case Report Forms). These data were then converted to SDTM (Study Data Tabulation Model) which would serve as the source for analysis datasets. However, real time availability of clinical data for analysis is becoming increasingly more important. Study teams need to continually look at clinical data to identify trends to ensure patient safety. In this paper we explore the feasibility of producing ADaM (Analysis Data Model) Datasets and analysis/analytics directly from FHIR Resources without having to transform the data to SDTM first. This allows users to layer analytics directly on top of FHIR resources pulled from an API in near real time. The benefit would be improved access to real time safety dashboards/signals and improved patient safety.

INTRODUCTION

In two previous papers, the PhUSE working group "Investigating the Use of FHIR in Clinical research" demonstrated that data typically collected in diabetes studies can be extracted from medical records through FHIR resources and the process can be automated to populate eCRFs. In this project we examine extending the use of FHIR to provide near real time analytics directly from EHRs (electronic healthcare records) and whether FHIR provides standardized data that can be used to drive the production of analytics. We develop a model and automated process to extract data from EHRs using FHIR resources and APIs (application programming interface), use these data as input for creating CDISC ADaM analysis datasets, which in turn are used to generate analytics/reports. This project builds on the previous paper (Hume, Abolafia, & Low, 2018) presented by the Research on FHIR (ROF) working group that demonstrated how the FHIR API along with the new CDISC ODMv2 standard can automate extracting data from EHRs and serving this data up for downstream processes.

Quick availability of clinical data for analysis is becoming more and more important in drug development. Subject safety is the most significant aspect in any drug exposure study. The medical and clinical operations teams need to look at the safety data to identify trends, adverse events, drug toxicity etc., as early as possible to ensure patient safety. That being the case, data, relevant safety dashboards, and analytics must be readily available to the study teams for review. However, the time from data collection, whether on paper CRFs or direct entry into an EHR, until entry in a study's EDC, is often two to six weeks. This is a significant hurdle to early identification of safety signals. In cases where source data are entered directly into EHRs this barrier can be overcome by generating analysis and alerts directly from the electronic medical data. Other use cases include long-term follow-up of patients and post marketing studies, where a new drug is rolled out into a broader population. In this scenario, EHR data can be continuously mined to monitor lab, vital signs, medication, and medical conditions at both an individual and aggregate level. At this point in time the Adverse Events (AE) FHIR resource is too immature to effectively monitor AEs, but it shows potential for future use.

BACKGROUND

CDISC ANALYSIS DATA MODEL (ADaM)

The CDISC ADaM standard defines the fundamental principles and standards to follow when creating analysis datasets and associated metadata (CDISC, 2018a). While the CDISC CDASH and SDTM standards provide

PhUSE EU Connect 2019

models for the collected data, the ADaM model specifies data and metadata structures to support the efficient generation and replication of analyses. The latest version of the ADaM Implementation Guide, Version 1.1, was released in February 2016.

CDISC SHARE

CDISC SHARE is a cloud-based CDISC standards metadata repository used to curate, manage, and publish standards metadata in machine-readable formats (Hume et al., 2018). CDISC SHARE is a curated resource that simplifies the retrieval and implementation of CDISC standards metadata in clinical research information systems such as clinical data management systems, mobile apps, and learning health systems (CDISC, 2018a). In addition to providing a resource to software developers, data managers, and biostatisticians, CDISC SHARE also supports the standards development process by enabling the implementation of new software tools and quality improvement mechanisms.

HL7 FAST HEALTHCARE INTEROPERABILITY RESOURCES (FHIR)

HL7 published FHIR Release 3 in March 2017 (HL7, 2018) as a standard for exchanging healthcare information electronically using a RESTful (Representational State Transfer) API and data organized into Resources. Resource content provides a focused, independent block of information that stands alone, or when combined with other Resources forms a useful patient record. Resources are new with FHIR, but their development has been informed by the previous generations of HL7 standards. FHIR seeks to broaden the scope of data sharing across organizations, disciplines, devices, and platforms, as well as enabling developers to integrate data faster than previously possible. FHIR aims to maintain relative simplicity by addressing the most common 80% of healthcare data exchange needs, while leaving the other 20% to extensions. FHIR has been developed based on today's web technologies to maintain a low learning curve, leverage existing software libraries, and maximize implementation feasibility.

eSOURCE

The FDA defines electronic source data, or eSource, as data initially recorded in electronic format, and EHR systems may function as an eSource data originator (FDA, 2013). The EHR functions as an electronic source when software transfers EHR data into an EDC system's CRF (FDA, 2013). The term eSource often refers to the pre-population of EDC-based CRFs using EHR (Electronic Data Capture) data (Mitchel, 2015). The FDA claims eSource benefits may include: eliminating unnecessary data duplication, reducing transcription errors, capturing source data at the time of the patient visit, promoting remote monitoring of data, encouraging real-time data reviews, and improving the accuracy and completeness of the data (FDA, 2013). The TransCelerate BioPharma eSource Initiative survey showed near consensus on the importance of clinical data standards in realizing EHR eSource integration for clinical research (Kellar et al., 2016).

REAL WORLD EVIDENCE (RWE)

RWE can come from a variety of sources including EHRs, payer administration claims, and patient registries (Sherman, 2016). One of the challenges in obtaining RWE from existing data sources stems from the diverse array of applications used to collect and store healthcare data. The HL7 FHIR standard can serve as an interface between diverse EHR applications delivering real world data (RWD) in support of RWE. The FDA currently uses RWE to monitor post-market safety and adverse events. The FDA has also indicated that data derived from RWE, under the right circumstances, can be used to support a marketing application. Furthermore, the 21st Century Cures Act, passed in 2016, emphasizes the use of RWE to support regulatory decisions.

SYNTHETICMASS DATASET

The SyntheticMass simulated Health Information Exchange (HIE) open-source dataset populated with EHR data for over 1 million simulated residents of Massachusetts statistically mirrors the real population in terms of demographics, disease burden, vaccinations, medical visits, and social determinants (Zopf et al., 2017). The Synthea synthetic patient population simulator generated the synthetic patients within SyntheticMass dataset. Synthea generates synthetic (not real), but realistic patient data and associated health records in several formats, including HL7 FHIR 3.0.1 (MITRE, 2017).

HEALTHCARE SERVICES PLATFORM CONSORTIUM SANDBOX

In 2013 Intermountain Healthcare, Louisiana State University, and the U.S. Department of Veteran Affairs founded the Healthcare Services Platform Consortium (HSPC) to refocus how healthcare applications are developed. Today over 270 contributors have joined the consortium to help develop a healthcare services platform community that supports the development of interoperable applications existing in a service-oriented architecture and knowledge-enabled model. HSPC's mission is to improve patient health by creating an "open ecosystem of interoperable

PhUSE EU Connect 2019

applications, knowledge, content, and services” (HSPC, 2017). The HSPC platform enables developers to create sandboxes consisting of an isolated environment, a partial SyntheticMass dataset, and software tools that support the development of healthcare applications.

METHODS

MAPPING FROM THE FHIR STANDARD TO ADaM

This project used the same database as in our two previous papers: adult subjects (greater than 18 years of age) with Diabetes Mellitus Type 2 with at least two encounters in the SyntheticMass Database. We extracted data points that included patient level medical records such as demographics, encounter information, medications, medical history, adverse events, laboratory and vital signs. The data and extraction process is described in more detail in our first paper (Zopf et al., 2017).

A key objective of this project was to map data from the FHIR standard to the CDISC ADaM standard. This was done so we could have standardized datasets that are analysis ready and contain standard derived variables. Another objective was to assess the degree of harmonization between FHIR and the ADaM CDISC standard. That is, the feasibility of easily mapping data that is in FHIR format to ADaM and whether we can develop standard scripts or programs that create ADaM data sets from EHR data. The FHIR and ADaM CDISC standards organize data or information differently. In ADaM, information is grouped into a series of datasets that are organized to facilitate analysis and reporting. ADaM datasets include data from multiple operational datasets, contain collected data as well as derived data, and are often restructured to be analysis ready.

In contrast, FHIR information is grouped into a collection of “Resources”, each containing a type of clinical or administrative information, such as prescriptions or procedures. Effectively, resources represent clinical “things” that contain small, logically discrete units of information that can be requested using the FHIR RESTful API. Resources have a defined meaning and represent the smallest “unit of transaction” of interest to healthcare.

The diagram in Figure 1 shows the high-level process followed by this project. The remaining sub-sections expand on the steps identified in each box included in the Figure 1 process diagram.

RESEARCH ON FHIR: USING EHR DATA AND CDISC ADAM FOR SAFETY ANALYSIS



Figure 1. High-level Process Diagram for Safety Analysis using CDISC ADaM and HL7 FHIR

Retrieve Patient Data Using the FHIR API

This project uses the HSPC sandbox (<https://sandbox.hspconsortium.org/rofphir/appsData>) as the EHR system with Synthea simulated patient data. The HSPC sandbox implements HL7 FHIR STU3 (Standard for Trial Use). The HSPC sandbox provides a FHIR API that enables client software to work with FHIR resources. For this project, we developed FHIR client software in Python to request specific clinical information from specified patients. FHIR clients could also be developed in R, SAS, Java, C#, and other popular languages. The FHIR API returns EHR data as resources structured according to the FHIR resource specifications listed on <http://hl7.org/fhir/STU3/index.html>. FHIR resources have explicit and stable URIs used to drive resource-based data exchange. Examples of FHIR resources used in this project include: Patient, Condition, Medication Statement, Encounter, and Observations. Observation resources, such as those for laboratory data or vital signs, are requested by providing the associated LOINC (Logical Observation Identifiers Names and Codes) code. A given LOINC code provides a description of the parameter, the unit of measurement, the specimen type, the method of the test and the category for the test or vital sign. The FHIR API returns resources represented using the JSON (Java Script Object Notation) media type, a widely used format for API-based data transfer. The screen shot shown in Figure 2 shows a template of the Patient FHIR resource in the JSON format.

Structure

UML

XML

JSON

Turtle

R3 Diff

All

JSON Template

```

{
  "resourceType" : "Patient",
  // from Resource: id, meta, implicitRules, and language
  // from DomainResource: text, contained, extension, and modifierExtension
  "identifier" : [{ Identifier }], // An identifier for this patient
  "active" : <boolean>, // Whether this patient's record is in active use
  "name" : [{ HumanName }], // A name associated with the patient
  "telecom" : [{ ContactPoint }], // A contact detail for the individual
  "gender" : "<code>", // male | female | other | unknown
  "birthDate" : "<date>", // The date of birth for the individual
  // deceased[x]: Indicates if the individual is deceased or not. One of these 2:
  "deceasedBoolean" : <boolean>,
  "deceasedDateTime" : "<dateTime>",
  "address" : [{ Address }], // An address for the individual
  "maritalStatus" : { CodeableConcept }, // Marital (civil) status of a patient
  // multipleBirth[x]: Whether patient is part of a multiple birth. One of these 2:
  "multipleBirthBoolean" : <boolean>,
  "multipleBirthInteger" : <integer>,
  "photo" : [{ Attachment }], // Image of the patient
  "contact" : [{ // A contact party (e.g. guardian, partner, friend) for the patient
    "relationship" : [{ CodeableConcept }], // The kind of relationship
    "name" : { HumanName }, // A name associated with the contact person
    "telecom" : [{ ContactPoint }], // A contact detail for the person
    "address" : { Address }, // Address for the contact person
    "gender" : "<code>", // male | female | other | unknown
    "organization" : { Reference(Organization) }, // C? Organization that is associated with the contact
    "period" : { Period } // The period during which this contact person or organization is valid to be c
    ontacted relating to this patient
  }],
  "communication" : [{ // A language which may be used to communicate with the patient about his or her h
    ealth
    "language" : { CodeableConcept }, // R! The language which can be used to communicate with the patie
    nt about his or her health
    "preferred" : <boolean> // Language preference indicator
  }],
  "generalPractitioner" : [{ Reference(Organization|Practitioner|
    PractitionerRole) }], // Patient's nominated primary care provider
  "managingOrganization" : { Reference(Organization) }, // Organization that is the custodian of the pati
    ent record
}

```

Figure 2. JSON Template for the Patient FHIR Resource

Convert FHIR JSON to ADaM SAS Datasets

The client software created in Python for this project converted the JSON FHIR resource content to CSV files for easy conversion by SAS into ADaM SAS datasets capable of producing data displays for typical safety domains: labs, vital signs, and adverse events. The following data were used for creating analysis data: demographics information (patients), medications, encounter information, description of conditions and laboratory and vital signs information (observations). As a result, we created four analysis data sets: ADSL (subject level data set), ADLB (lab analysis dataset), ADVS (vital signs analysis dataset), and ADAE (adverse events analysis dataset.). The ADaM datasets were limited in scope to just include variables needed for our displays.

We created an ADaM mapping specification document to describe how each ADaM dataset and variable was created from the FHIR formatted input data. We performed the mapping manually, since ADaM datasets often contain a host of complex derived variables. Table 1 displays the ADaM mapping specification.

The first ADaM dataset we mapped was ADSL, the subject level dataset. ADSL contains unique subject ID, study identifier, the subject's treatment group, and the start and end dates for the study drug. Unique study ID was created by de-identifying the identifier obtained from the subject's EHR. Study identifier did not exist in the EHR, so we just

PhUSE EU Connect 2019

assigned it a constant value. Treatment group also was not contained in the source data. For the purposes of our prototype, we used the subject's first visit in the database. We decided that Metformin Hydrochloride would be our study drug. If a subject was taking Metformin Hydrochloride at their first visit, we assigned the subject to the Metformin Hydrochloride treatment group. Otherwise, the subject was assigned to the Comparator group. The start and end dates associated with the treatment mapped directly from the FHIR formatted EHR data to their ADaM counterparts. All variables in the ADSL dataset were merged onto all other ADaM datasets.

ADLB, the ADaM analysis datasets for lab data, was the second data set we mapped. In addition to the variables in ADSL, ADLB contained the name of the lab parameter, the value of the parameter, the date and visit the measure was taken, normal lower and upper limits for a given parameter, the baseline value and the change from baseline value for a parameter. Since the source EHR data did not contain visit, we created a visit variable. We set a subject's first patient encounter in the EHR database to Visit 1. This record would also serve as the baseline record. Each additional patient encounter was assigned a sequential visit number, so that visit mirrored the subject's encounter number.

Laboratory parameters are specified in the Synthea FHIR data by a LOINC code (*coding.code*) and an associated text description (*coding.display*). A given LOINC code provides a description of the parameter, the unit of measurement, the specimen type, the method of the test and the category for the test or vital sign. Lab parameters and their associated values were extracted from these LOINC codes (for a more extensive discussion of this topic, see our first paper referenced above). At this point in time LOINC codes are not harmonized with CDISC controlled terminology, so this was a manual and somewhat tedious process. Finally, normal upper and lower limits for each lab parameter were assigned based on published values. The mapping from FHIR to ADaM is exactly the same for Vital Signs as for Laboratory data and therefore is not discussed here.

Table 1. Mapping Specification: FHIR to CDISC ADaM

Dataset	ADaM Variable	FHIR Resource	FHIR Field	Notes
ADSL	USUBJID	Patient	Identifier.Value	de-identify
	STUDYID			Assign (= 'CDISC01')
	TRT01A	Medication	coding.code/coding.display	Actual treatment 'Metformin Hydrochloride' or 'Comparator'
	TRTSDT	Medication Statement	effectivePeriod.start	
	TRTEDT	Medication Statement	effectivePeriod.end	
ADLB/ ADVS	USUBJID	Patient		ADSL.USUBJID
	STUDYID			ADSL.STUDYID
	TRTA			ADSL.TRT01A
	ADT	Observation	effectiveDateTime	date measure taken
	AVISITN			patient encounter number for each subject
	PARAM	Observation	coding.code	extract from LOINC code
	AVAL	Observation	valueQuantity.value	
	BASE			value of PARAM from first subject visit
	ABLFN			set to 1 if first patient visit
	CHG			AVAL - BASE
	ANRLO			assign from published standards
	ANRHI			assign from published standards
ADAE	USUBJID			ADSL.USUBJID

PhUSE EU Connect 2019

	TRTA			ADSL.TRT01A
	ASTDT	Conditions	conditions.start	start date of AE
	AENDT	Conditions	conditions.stop	end date of AE
	AETERM	Conditions	conditions.description	reported term or condition
	AESER			if encounter.code in (183478001,50849002,308646001) then AESER = Y else AESER =N
	AEENRF			if AENDT is missing then set to ONGOING
	AESEV			if encounter.code in (183478001,50849002,308646001) then aesev ="Severe". If encounter.code in (305411003,305408004) then aesev="Moderate" else "Mild"
	TRTEMFL			if astdt >= trtsdt then trtemfl="Y" else trtemfl="N"

The adverse event analysis dataset ADAE was the final dataset we mapped. As in ADLB we included all variables from ADSL, as well as, the “reported term” for the adverse event, the start and end dates associated with the adverse event, the severity of the adverse event, indicators for whether the adverse event was serious or ongoing, and a derived variable for whether the adverse event was treatment emergent. The FHIR condition resource was used to extract “adverse events”. This resource is still relatively immature and we likely did not capture all of what are traditionally termed adverse events in a clinical trial. For our prototype, we only used the reported term and made no attempt to code or recode the term. In our FHIR data conditions/adverse events were coded in SNOMEDCT (Systemized Nomenclature of Medicine – Clinical Terms), not MEDRA (Medical Dictionary for Regulated Activities) which is usually used in clinical research. It is important to note that the Condition resource contains a field called *coding.system* that links to the coding dictionary used. As a result, FHIR is not limited to a single coding system. Adverse Event severity and seriousness were not present in our EHR data, so for demonstration purposes, we derived them. If a subject had emergency room admission, death or emergency encounter we set AESER (indicator for seriousness) to “Y” and AESEV (severity) to ‘severe’. The final concept we mapped was a treatment emergent flag. If an adverse event occurred between treatment start date and end date, we considered the adverse event treatment emergent.

Table 2. Snapshot of ADSL Data

Study Identifier	Unique Subject Identifier	Date of First Exposure to Treatment	Date of Last Exposure to Treatment	Actual Treatment for Period 01
CDISC01	c0d2009f-a2e2-4362-9027-f8e2aa38fb6a	1977-04-18		Metformin Hydrochloride
CDISC01	c228478d-5512-4e3b-9bb6-6c315ba7a74a	2005-03-09		Metformin Hydrochloride
CDISC01	c7d4923d-462c-43ec-8c4d-94f17e1fd9cd	1984-09-17		Metformin Hydrochloride

PhUSE EU Connect 2019

Table 3. Snapshot of Lab Analysis Datasets (ADLB)

Study Identifier	Unique Subject Identifier	Date of First Exposure to Treatment	Date of Last Exposure to Treatment	Actual Treatment	Analysis Date
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	1997-06-04
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	1998-06-10
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	1998-11-25
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	1999-03-24
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	1999-05-26
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2000-01-19
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2000-12-13
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2001-08-08
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2001-11-07
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2002-04-17
CDISC01	054cfd5b-ccb0-4cb0-baf2-eacea220545f	1992-08-19		Metformin Hydrochloride	2002-07-10

ADLB continued:

Parameter	Analysis Value	Baseline Record Flag	Baseline Value	Analysis Visit	Analysis Visit (N)	Change from Baseline	Analysis Normal Range Lower Limit	Analysis Normal Range Upper Limit
Glucose (mg/dL)	141.8	Y	141.8	Baseline	1	0	65	110
Glucose (mg/dL)	159.5		141.8	VISIT 2	2	17.7	65	110
Glucose (mg/dL)	192.7		141.8	VISIT 3	3	50.9	65	110
Glucose (mg/dL)	137.6		141.8	VISIT 4	4	-4.2	65	110
Glucose (mg/dL)	160.5		141.8	VISIT 5	5	18.7	65	110
Glucose (mg/dL)	180.7		141.8	VISIT 6	6	38.9	65	110
Glucose (mg/dL)	141.8		141.8	VISIT 7	7	0	65	110
Glucose (mg/dL)	135.8		141.8	VISIT 8	8	-6	65	110
Glucose (mg/dL)	156.6		141.8	VISIT 9	9	14.8	65	110
Glucose (mg/dL)	199		141.8	VISIT 10	10	57.2	65	110
Glucose (mg/dL)	149.3		141.8	VISIT 11	11	7.5	65	110
Glucose (mg/dL)	186.7		141.8	VISIT 12	12	44.9	65	110

Table 4. Snapshot of the Adverse Events Analysis Dataset (ADAE)

Study Identifier	Unique Subject Identifier	Date of First Exposure to Treatment	Date of Last Exposure to Treatment	Actual Treatment
CDISC01	1e94a0ae-1645-488d-bd9e-d87625aac1d0	1978-04-19		Metformin Hydrochloride
CDISC01	1e94a0ae-1645-488d-bd9e-d87625aac1d0	1978-04-19		Metformin Hydrochloride
CDISC01	1e94a0ae-1645-488d-bd9e-d87625aac1d0	1978-04-19		Metformin Hydrochloride
CDISC01	1e94a0ae-1645-488d-bd9e-d87625aac1d0	1978-04-19		Metformin Hydrochloride
CDISC01	2061fa0e-237b-4b23-b571-bdfe04976d08	1981-12-05		Metformin Hydrochloride
CDISC01	2061fa0e-237b-4b23-b571-bdfe04976d08	1981-12-05		Metformin Hydrochloride
CDISC01	2061fa0e-237b-4b23-b571-bdfe04976d08	1981-12-05		Metformin Hydrochloride
CDISC01	2061fa0e-237b-4b23-b571-bdfe04976d08	1981-12-05		Metformin Hydrochloride

PhUSE EU Connect 2019

ADAE continued:

Reported Term for the Adverse Event	Analysis Start Date	Analysis End Date	End Relative to Reference Period	Serious Event	Severity/ Intensity	Treatment Emergent Analysis Flag
Acute viral pharyngitis (disorder)	2007-03-05	2007-03-15		N	Mild	Y
Viral sinusitis (disorder)	2001-08-19	2001-08-26		N	Mild	Y
Viral sinusitis (disorder)	2002-07-31	2002-08-07		N	Mild	Y
Viral sinusitis (disorder)	2003-05-09	2003-05-23		N	Mild	Y
Acute viral pharyngitis (disorder)	1997-03-22	1997-04-03		N	Mild	Y
Acute viral pharyngitis (disorder)	2004-10-27	2004-11-09		N	Mild	Y
Seizure disorder	1971-11-22		ONGOING	Y	Severe	N
Sinusitis (disorder)	1999-03-04	1999-04-08		N	Mild	Y

Execute Analysis Programs

For the purposes of this prototype, SAS was used to create ADaM datasets, summary tables and graphical displays. A batch file to run the analysis programs can be created in SAS, R, or Python as needed. APIs can be used to run the batch file and display the tables and graphs in real-time. For this prototype, a SAS batch file was created to refresh the ADaM datasets and execute the programs to create the tables and figures. Proc HTTP can be used to post the displays to a web browser. SAS code for a POST request will look like the following:

```
1. filename in "input file name";
2. filename out "output file name";
3.
4. proc http
5. url = "http://httpbin.org"
6. in = in
7. out =out;
8. run;
```

Display Safety Analysis Results

In this prototype we generate displays showing change from baseline for laboratory data as well as treatment emergent adverse events at the patient and treatment group levels. When our script updates the ADaM datasets, displays will also be refreshed. These displays are easily reproducible from standard ADaM datasets.

RESULTS

We successfully mapped all variables needed for our analyses from FHIR to ADaM with a few caveats. To create the "treatment" variable, if a subject was taking Metformin Hydrochloride at their first encounter in the database, we assigned the subject to the 'Metformin Hydrochloride' treatment group. Otherwise we assigned the subject to the 'Comparator' treatment group. We assumed treatment group remained constant for subsequent subject encounters. We used patient encounter number as a proxy for visit number. In contrast to a research protocol, patient encounters did not occur at regular pre-defined intervals and the time between subject encounters differed for each subject. For ADAE, while severity and seriousness of an adverse event are concepts that exist in FHIR, these fields were not available in our sample data. We derived them as described above. Also, we did not recode the reported AE term, which was coded using SNOMEDCT. This is consistent with most EHR systems and highlights the issue that in many cases there is a lack of harmonization between EHR/FHIR and CDISC/research data.

Data was successfully retrieved from the Synthea database and used as input to create ADaM analysis datasets specified in the mapping documents. Our script then fired off the SAS programs that generated a safety dashboard. Below we describe some of the analyses that may be generated from the dashboard as well as some scenarios for their use.

PhUSE EU Connect 2019

Since our therapeutic area of interest was Diabetes Mellitus Type 2, a medical monitor or a patient's physician was first interested in monitoring a subject's glucose and adverse events over time. **Figures 3 and 4** list subject level glucose data and adverse events while the subject was on study drug. After viewing these results, the medical monitor was

interested in how a selected patient compared to other patients at the site in regards to glucose levels over time and adverse events. **Figure 5** shows change from baselines from patients in the study drug group compared to the comparator group. **Figure 6** presents selected adverse events in our two treatment groups.

Figure 3. Change from Baseline Glucose for Patient XXX

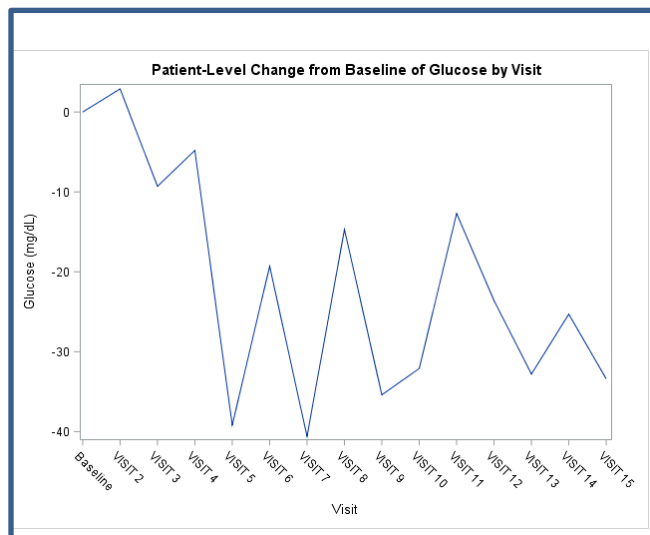


Figure 4. Treatment Emergent Adverse Events for the Patient XXX

Treatment Emergent Adverse Event					
Treatment: Metformin Hydrochloride					
Patient Number	Adverse Event	Start Date	End Date	Serious AE	Severity
1460e004-4c03-4244-b350-dbfd2270ed48	Acute viral pharyngitis (disorder)	13-APR-1986	24-APR-1986	N	Mild
	Injury of tendon of the rotator cuff of shoulder	09-MAY-1989	15-JUL-1989	N	Mild
	Normal pregnancy	30-NOV-1981	07-DEC-1981	N	Mild
	Viral sinusitis (disorder)	02-FEB-1981	09-FEB-1981	N	Mild

Figure 5. Mean Change from Baseline Glucose by Treatment Group

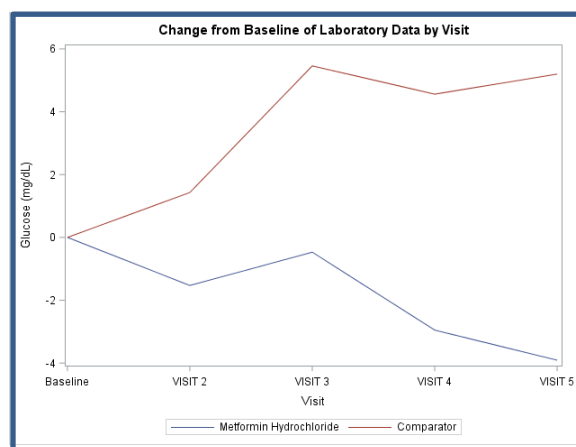
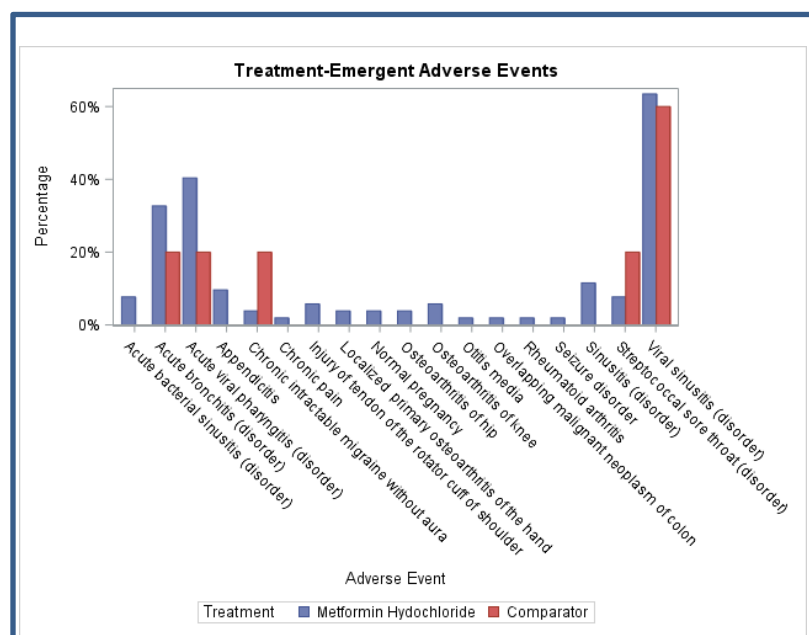


Figure 6. Treatment Emergent Adverse Events by Treatment Group



DISCUSSION

Retrieving EHR Patient Data using HL7 FHIR

Most RESTful APIs exchange data using JSON as the default media type, though APIs typically support multiple media types. JSON is language-independent file format used to transmit data objects across applications. JSON is a compact and lightweight data interchange format. It is an open-standard format used widely for transmitting data.

R and Python have capabilities to read in JSON files directly. In SAS, version 9.4 maintenance 4 onwards have capabilities to read in JSON files directly to create SAS datasets. When using SAS versions prior to 9.4, using a language like Python to convert the data from JSON to CSV makes the conversion simpler. Using the Python JSON and

PhUSE EU Connect 2019

CVS libraries makes simple work of converting FHIR JSON-based EHR data to CSV files. Once in CSV format, converting CSV to SAS datasets becomes a simple task in SAS.

The client software used to retrieve the FHIR EHR data in JSON and convert it to CSV files for later conversion to SAS datasets was developed using Python 3.6.2. To retrieve EHR data from the HSPC sandbox using the FHIR API, the client software used the Python SMART on FHIR *fhirclient* 3.0.0 library [ref <https://github.com/smart-on-fhir/client-py>]. The Python *fhirclient* library works with FHIR servers that support the SMART on FHIR protocol, such as the HSPC sandbox FHIR server. The HSPC sandbox with SyntheticMass data implements FHIR v3.0.1 and is openly accessible at <https://api-v5-stu3.hspconsortium.org/rofpfir/open>.

The Python *fhirclient* library not only provides pythonic ways to retrieve data directly or via searches, but also provides data model classes such that client developers can work with FHIR data as classes rather than working with the raw JSON returned by the server. For example, the code snippet below shows *fhirclient* using the *Patient* class to create a patient object. This example uses *fhirclient* to retrieve the latest Patient resource from the HSPC sandbox EHR server for a patient with the identifier "SMART-1288992". Using the *Patient* class allows developers to access gender, for example, as an attribute of the patient object using *patient.gender* instead of parsing the raw JSON directly to retrieve gender.

```
1. from fhirclient import client
2. import fhirclient.models.patient as p
3. smart = client.FHIRClient(settings=settings)
4. patient = p.Patient.read('SMART-1288992', smart.server)
5. print(patient.name[0].given[0], patient.name[0].family)
6. print(patient.birthDate.isostring)
7. print(patient.gender)
```

PROTOTYPE LIMITATIONS

Our prototype demonstrates an automated repeatable process that can be utilized to retrieve EHR data in near real time and provide safety analytics to a medical monitor or patient's physician. Our prototype makes several assumptions for demonstration purposes, including some that may be unrealistic. Since we used the Synthea database described above, we had access to all patient data. We did not have to deal with any security or access issues. Second, we only attempted to extract a small percentage of a patient's record. There are many additional variables an end user may be interested in such as additional medications, procedures, or events. It could be that these data are difficult to model in the FHIR standard or are not easily harmonized with the CDISC standard. Third, the FHIR Resources for extracting adverse events data are still immature. It's likely we did not capture all of a subject's adverse events data accurately. Fourth, we dealt with treatment or study drug in an oversimplified manner. We assigned treatment group based on a subject's first encounter in the EHR database. We then assumed treatment remained unchanged over time.

FHIR/CDISC HARMONIZATION

While we successfully mapped all of our variables of interest from FHIR to ADaM, it required more effort than if the input data format was CDASH or SDTM. The CDISC CDASH and SDTM standards are harmonized with the CDISC ADaM standard, while the FHIR standard is not. Many of the ADaM variables are exact copies of their CDASH/SDTM counterparts. Furthermore, controlled terminologies in FHIR and CDISC are not harmonized. For example, FHIR typically uses SNOMEDCT, while CDISC uses MEDRA. For our project and others with limited scope, which only mapped a small number of variables and domains, this is a manageable problem. For larger and more complex implementation this may not be the case.

CONCLUSION

This project builds on the previous two research projects by the Research on FHIR Team, constructing a prototype that generates near real time analytics from EHRs. The prototype demonstrated an automated repeatable process that: 1) used FHIR APIs to retrieve selected patient data from an EHR; 2) used a standard script to create CDISC ADaM datasets from the FHIR formatted EHR data; and 3) produced sample analytics that can be generated from ADaM datasets. Our prototype not only allows medical or clinical staff to get near real time information about a patient they are treating, but also gives them the ability to compare patient outcomes on a given treatment to others on the same or alternative treatments.

The project illustrates how existing standards for provider and research data can be integrated to provide more timely safety analytics to medical staff. It also demonstrates the lack of harmonization between provider (FHIR) and research

PhUSE EU Connect 2019

(CDISC) standards. In the future the pharmaceutical industry should consider moving to provider terminologies such as SNOMEDCT and LOINC. This would increase the interoperability between research and provider data and eliminate many error-prone and time consuming manual processes.

Future work should expand our prototype to multiple sites and EHR systems. This will increase the value and power of the analytics generated. As FHIR matures as a standard, improved mechanisms for capturing adverse events should be explored. Finally, our scope was limited. Future projects should expand the safety assessments we analyze and integrate EHR data with other relevant data sources.

REFERENCES

- CDISC. (2018a). CDISC SHARE. Retrieved from <https://www.cdisc.org/standards/share>
- CDISC. (2018b). Clinical Data Acquisition Standards Harmonization (CDASH). Retrieved from <https://www.cdisc.org/standards/foundational/cdash>
- CDISC. (2018c). Operational Data Model (ODM). Retrieved from <https://www.cdisc.org/standards/transport/odm>
- FDA. (2013). *Guidance for Industry: Electronic Source Data in Clinical Investigations*. FDA Retrieved from <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm328691.pdf>.
- HL7. (2018). FHIR Release 3. Retrieved from <http://hl7.org/fhir/STU3/index.html>
- HSPC. (2017). Healthcare Services Platform Consortium Sandbox. Retrieved from <https://sandbox.hspconsortium.org>
- Hume, S., Aerts, J., Sarnikar, S., & Huser, V. (2016). Current applications and future directions for the CDISC Operational Data Model standard: A methodological review. *Journal of Biomedical Informatics*. doi:doi:10.1016/j.jbi.2016.02.016
- Hume, S., Chow, A., Evans, J., Malfait, F., Chason, J., Wold, D., . . . Becnel, L. (2018, March 15, 2018). *CDISC SHARE, a Global, Cloud-based Resource of Machine-Readable CDISC Standards for Clinical and Translational Research*. Paper presented at the AMIA Informatics Summit 2018, San Francisco, CA.
- Hume, S., Abolafia, J., Low, G., (2018). *Use of HL7 FHIR as eSource to Pre-populate CDASH Case Report Forms Using a CDISC ODM API*. Paper presented at the PhUSE US Connect Conference 2018, Raleigh, NC.
- Kellar, E., Bornstein, S. M., Caban, A., Célingant, C., Crouthamel, M., Johnson, C., . . . Wilson, B. (2016). Optimizing the Use of Electronic Data Sources in Clinical Trials: The Landscape, Part 1. *Therapeutic Innovation & Regulatory Science*, 50(6), 682-696. doi:10.1177/2168479016670689
- Mitchel, J. T., Helfgott, J., Haag, T., Cappi, S., McCanless, I., Kim, Y.J., Choi, J., Cho, T., Gittleman, D.A. (2015). eSource Records in Clinical Research. *Applied Clinical Trials*.
- MITRE. (2017). SyntheticMass. Retrieved from <https://syntheticmass.mitre.org/index.html>
- S4S. (2018). Sync for Science. Retrieved from <http://syncfor.science/>
- Wiffin, E., Mandel, J., Schwertner, N., Pfiffner, P., & Baranika, T. (2018). Python SMART on FHIR client. github.com. Retrieved from <https://github.com/smart-on-fhir/client-py>
- Zopf, R., Abolafia, J., & Reddy, B. (2017). *Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE)*. Paper presented at the PhUSE Annual Conference 2017, Edinburgh, Scotland.

ACKNOWLEDGMENTS

The authors would like to thank Trisha Simpson (UCB), Mike Hamidi (CDISC), and Ann White (CDISC) for their contributions to the paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the team co-lead at:

Trisha Simpson
UCB
8010 Arco Corporate Drive
Suite 100
Raleigh, NC 27617
+ 919-767-3107
Trisha.Simpson@ucb.com

Brand and product names are trademarks of their respective companies.