Contemporary Clinical Trials Communications 28 (2022) 100920

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Using passive extraction of real-world data from eConsent, electronic patient reported outcomes (ePRO) and electronic health record (EHR) data loaded to an electronic data capture (EDC) system for a multi-center, prospective, observational study in diabetic patients

Cynthia M. Senerchia [a], Tracy L. Ohrt [a,*], Peter N. Payne [a], Samantha Cheng [a], David Wimmer [a], Irene Margolin-Katz [a], Devin Tian [a], Lawrence Garber [b], Stephanie Abbott [c], Brian Webster [d]

[a] Optum Insight, Life Sciences, Digital Research Network, 11000 Optum Circle, Eden Prairie, MN, 55344, USA
[b] Reliant Medical Group, 640 Lincoln Street, Worcester, MA, 01605, USA
[c] Western Washington Medical Group, 1728 W, Marine View Drive, Suite 110, Everett, WA, 98201, USA
[d] Wilmington Health, 1202 Medical Center Drive, Wilmington, NC, 28401, USA

ARTICLE INFO

Keywords:
Electronic source
Pragmatic clinical trial
Real world evidence
Real world data
Electronic health record
Electronic data capture

ABSTRACT

As clinical trial complexity has increased over the past decade, using electronic methods to simplify recruitment and data management have been investigated. In this study, the Optum Digital Research Network (DRN) has demonstrated the use of electronic source (eSource) data to ease subject identification, recruitment burden, and used data extracted from electronic health records (EHR) to load to an electronic data capture (EDC) system. This study utilized electronic Informed Consent, electronic patient reported outcomes (SF-12) and included three sites using 3 different EHR systems. Patients with type 2 diabetes with an HbA1c $\geq$ 7.0% treated with metformin monotherapy were recruited. Endpoints consisted of changes in HbA1c, medications, and quality of life measures over 12-weeks of study participation using data from the subjects' eSources listed above. The study began in June of 2020 and the last patient last visit occurred in January of 2021. Forty-eight participants were consented and enrolled. HbA1c was repeated for 33 and ePRO was obtained from all subjects at baseline and 28 at 12-week follow-up.

Using eSource data eliminated transcription errors. Medication changes, healthcare encounters and lab results were identified when they occurred in standard clinical practice from the EHR systems. Minimal data transformation and normalization was required.

Data for this observational trial where clinical outcomes are available using lab results, diagnoses, and encounters may be achieved via direct access to eSources. This methodology was successful and could be expanded for larger trials and will significantly reduce staff effort and exemplified clinical research as a care option.

1. Introduction

Clinical trials are considered the foremost authority for developing medical evidence to support safety and efficacy of new drugs, biologics, or devices due to their methodological strengths [1–3]. However, clinical trials often experience challenges including timely accrual, high

Abbreviations: AE, Adverse Event; DRN, Digital Research Network; eCRF, Electronic Case Report Form; EDC, Electronic Data Capture; EHR, Electronic Health Record; ePRO, Electronic Patient Reported Outcomes; HbA1c, Hemoglobin A1C; HCO, Healthcare Organization; HIPAA, Health Insurance Portability and Accountability Act; ICD-10, International Classification of Diseases 10th Revision; IDN, Integrated Delivery Network; IP, Investigational Product; IQR, Interquartile Range; IRB, Institutional Review Board; QOL, Quality of Life; RWD, Real World Data; RWE, Real World Evidence; SAE, Serious Adverse Event; SD, Standard Deviation; SDV, Source Data Verfication; SF-12, Short Form Health Survey; SMS, Optum Smart Measurement System; T2DM, Type 2 Diabetes Mellitus.

* Corresponding author.
E-mail addresses: cynthia.senerchia@optum.com (C.M. Senerchia), tracy.ohrt@optum.com (T.L. Ohrt), peter.payne@optum.com (P.N. Payne), samantha.cheng@optum.com (S. Cheng), david.wimmer@optum.com (D. Wimmer), irene.margolin-katz@optum.com (I. Margolin-Katz), devin.tian@optum.com (D. Tian), Lawrence.Garber@reliantmedicalgroup.org (L. Garber), sabbott@wwmedgroup.com (S. Abbott), BWebster@wilmingtonhealth.com (B. Webster).

https://doi.org/10.1016/j.conctc.2022.100920
Received 8 September 2021; Received in revised form 17 March 2022; Accepted 1 May 2022
Available online 5 May 2022
2451-8654/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

operational costs, or extensive and stringent data requirements [4,5]. These complexities result in most studies being conducted at a small number of sites across the United States (US). Therefore, findings may have limited generalizability because the eligibility criteria result in a study population that is not representative of patients with the disease. These settings may result in limited participation by minorities, as these sites are not where all health care is provided within the US [6]. These concerns support investigation of new trial methods. Real-world data (RWD), including data collected within electronic health records (EHRs), may provide a solution. These datasets contain diverse patient populations with data collected often over a prolonged period [7].

To address these concerns, this pragmatic study enrolled participants with type 2 diabetes mellitus (T2DM), the most common diabetes in adults, resulting in deleterious health outcomes when not controlled [8]. Adults with inadequately controlled T2DM (defined as HbA1c $\geq$ 7%) being treated with metformin monotherapy were eligible. The intent of this study was to gather insight into the feasibility of using EHR and electronic patient reported outcomes (ePRO) surveys as Real-world Evidence (RWE). The target population was expected to experience at least one healthcare encounter, medication adjustments, and additional HbA1c measurements during the planned 12-week follow-up. The objectives of this Proof of Concept study were 1) to confirm EHR data was appropriate for determining the study endpoints, was accessible and of suitable quality to support its use as RWE; as well as determine the effort required to gather these data and 2) use these data to observe the impact on participants' clinical care parameters such as: changes in HbA1c, antidiabetic medications and self-reported quality of life (QoL) in a real-world setting.

2. Methods

Optum, as a business associate of many healthcare organizations (HCO), has access to near real-time EHR data for approximately 104 million patients across the US, and approximately 15 million patients receive care at HCOs in the Digital Research Network (DRN). These HCOs represent a diverse patient population from academic medical centers with integrated delivery networks (IDNs) and community-based ambulatory care settings. This prospective, observational study was conducted at 3 large HCOs that together manage over 2 million patients from different regions of the US: New England, Northwest, Southeast. All were experienced in research and use different EHR systems. There were no scheduled visits, and no investigational product (IP) was administered. Potential participants were identified by applying the eligibility criteria to the site's EHR data using a coding algorithm. Participants who provided eConsent were considered enrolled, initiating the passive collection of data ambispectively from their EHR data collected during routine healthcare interactions. An ePRO, Short Form Health Survey (SF-12), was used to assess any changes in the participants' self-reported QoL over the study period [9,10]. The electronic source (eSource) data were copied to a stage database, normalized, transformed, loaded to the study database, and subsequently extracted and transferred to the electronic data capture (EDC) system using the processes described below (see Fig. 1).

2.1. Technical methods

Prior to study initiation, Optum evaluated each EHR system to determine the source table for the required data elements. A comprehensive EHR data mapping and transformation specification was created to define the source data location; the normalization or transformation that would be required; and the 'landing' location of those data in the stage and study databases and the EDC. After initial analysis, Optum worked closely with the sites' subject matter experts to confirm the source to target data mappings. Finally, Optum extracted test data to confirm the study data requirements could be met. Optum confirmed that all required study data was present and available in the EHR systems as described below. Fig. 2 provides a pictorial view of the mapping process from EHR to study database.

The eConsent and ePRO vendors sent sample files for testing; and Optum verified that the data within the eConsent allowed participants to be found in the EHR. Following completion of verification and validation testing to confirm only consented participants' data was extracted, Optum assigned a unique Study Subject ID to each subject's EHR 'identity' using a subject map table. As data were entered by the site into the EHR, the data were flagged by the system as relevant, copied, and loaded to the stage environment. The normalization processes were executed as the data was transferred from the stage to the study database (Fig. 3). Source data as well as the normalized data and metadata were stored in the study database. The sequence of data collection is summarized in the schedule of events in the results section (Table 1).

The most recent EHR data at the time of consent were used to confirm eligibility and gather baseline data. On a daily basis, any new or changed data received for the participants was reviewed to confirm that the data was in the correct format, there were no errors in the extraction,

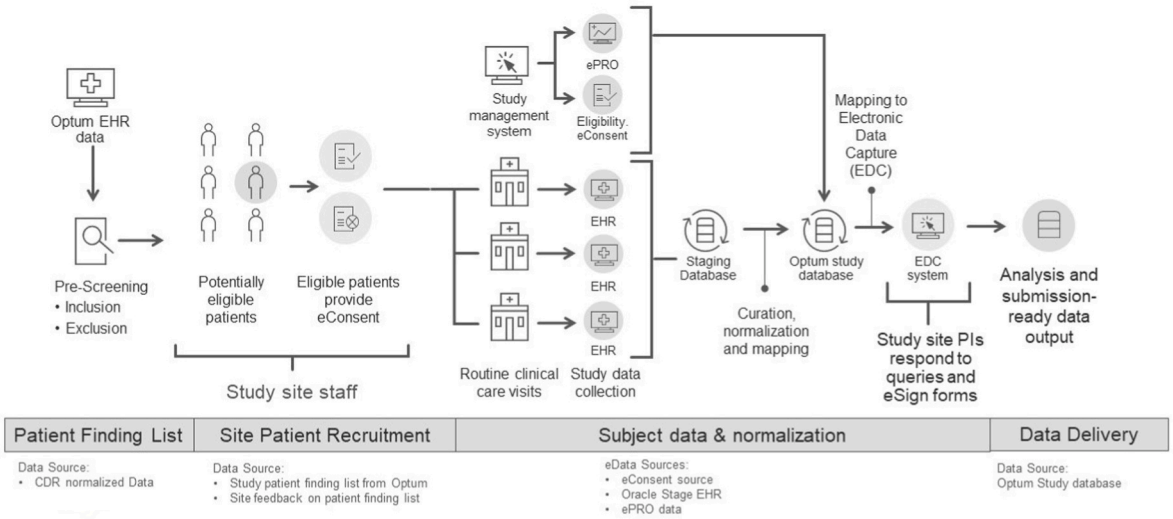

Fig. 1. Diagram of the end-to-end data flow
CDR = Common Data Repository; EHR = electronic health record; ePRO = electronic patient-reported outcome; EDC: Electronic data capture; PI: Principal Investigator.

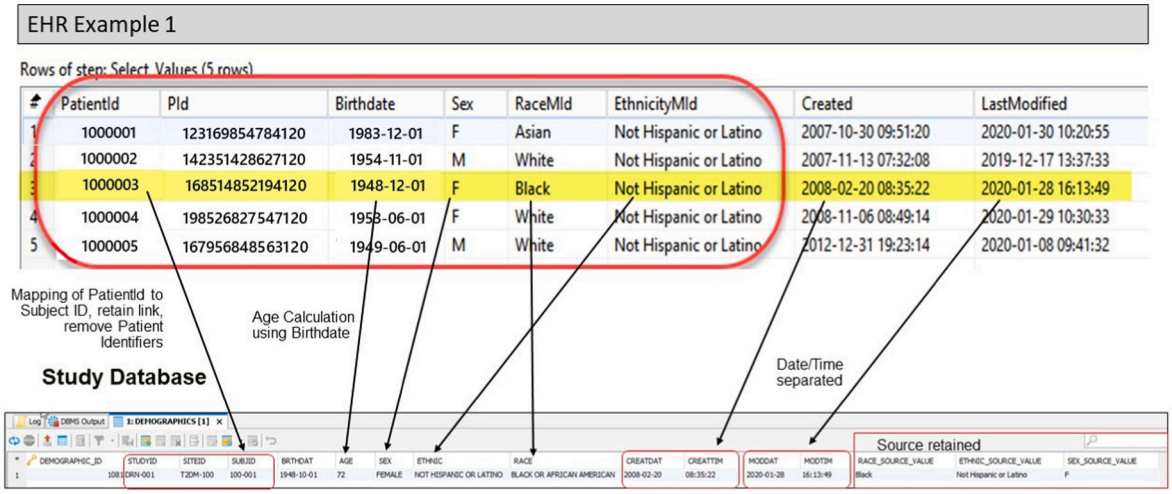

Fig. 2. Data mapping EHR to study database.

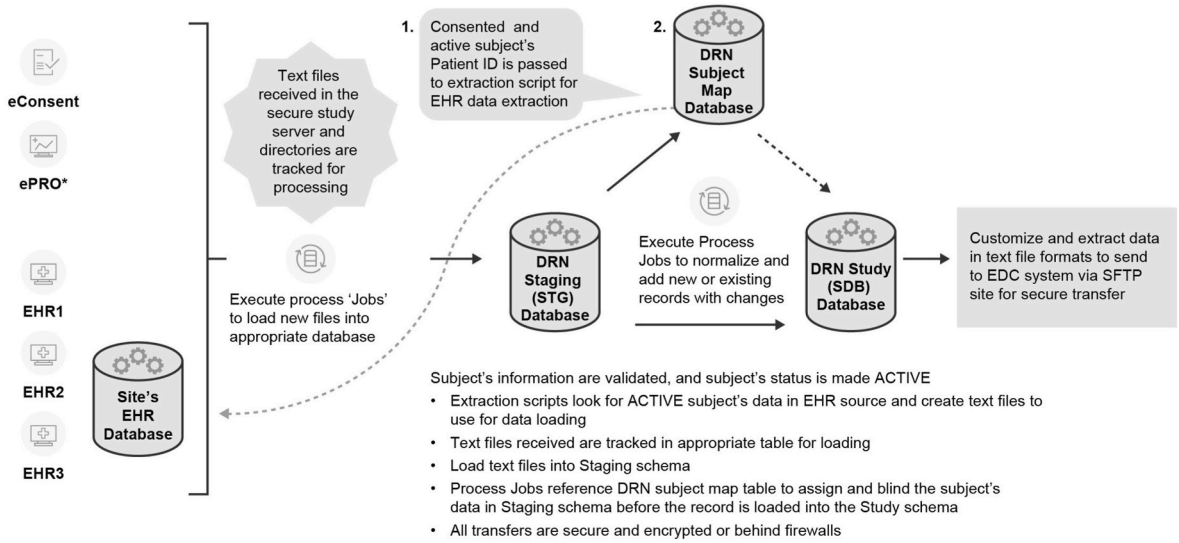

Fig. 3. Clinical study data Acquisition and normalization.

the data was correctly transformed or normalized and met the requirements for loading into the study database and the EDC system. Following these checks, data was transferred to the EDC vendor (Adaptive Clinical) for loading into the system built for the study (Fig. 4).

The eCRFs within the EDC system were developed to maintain both the *original source values,* the *normalized data*, as well as *the EHR metadata* (original date/time of entry, and the date/time the data was last updated in the EHR). The site staff were able to log into the EDC system to review

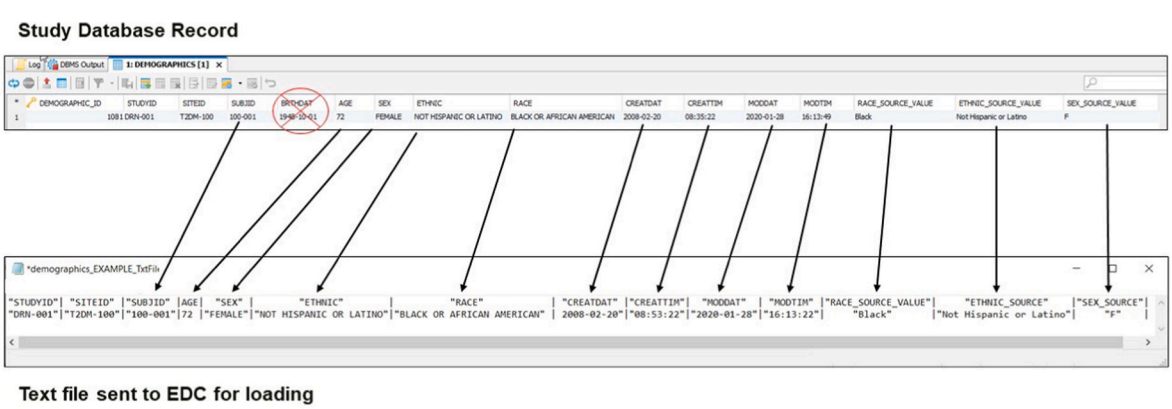

Fig. 4. Study database map to text file for EDC

the data to confirm correctness of the extraction and normalization processes, but no changes were allowed in the EDC system – any corrections occurred in the EHR.

3. Data quality assurance

Optum implemented and verified several data quality processes for error identification and query generation prior to study initiation. During the study, Optum reviewed the data files prior to forwarding files to the EDC system, and confirmed that each file loaded successfully. Optum verified that updates to EHR data was loaded without overwriting the original entry, and a new EDC audit record was created. Test records and objective evidence were maintained in the Optum electronic test systems for archival and audit purposes. Methods to verify the data was attributable, legible, contemporaneous, original, accurate and complete at study completion are detailed below.

3.1. Study initiation

The study enrolled adults ≥18 years with T2DM currently treated with metformin as monotherapy. The inclusion criteria were: 1) ≥18 years, 2) clinically confirmed T2DM (based on ≥2 ICD-10 diagnosis codes from ≥2 EHR encounters), 3) HbA1c results ≥7.0% (based on most recent laboratory value from EHR within the previous 12 months), 4) currently on metformin or metformin extended release, 5) capable and willing to provide eConsent, 6) capable and willing to complete ePRO. The exclusion criteria were: 1) On anti-diabetic medication other than metformin or in combination with metformin within 6 months prior to study enrollment, 2) currently participating in an interventional trial, or have participated in one in the last 30 days.

Optum identified potentially eligible participants by applying an algorithm to the most current data available. Lists of potentially eligible participants were populated into a pre-screening log including patient identifiers, contact information and confirmation of eligibility based on the most recent HbA1c and medications. In addition, the name of the participant's provider and next scheduled appointments were included to assist the research staff's outreach to participants. The lists were securely transferred to each site.

After confirming eligibility of the participants provided on the list, the staff approached the participants to enroll in the study. Each participant reviewed was assigned one of the following statuses for tracking: 1) consented/enrolled, 2) ineligible, 3) not a good candidate, 4) approached, not interested, 5) pending eligibility confirmation, 6) confirmed eligible in EHR, 7) follow-up needed. Potential participants with a status of consented/enrolled, ineligible, not a good candidate, or approached but not interested were removed from subsequent lists, and new potentially eligible participants were added.

The protocol, amendments, Informed Consent Form (ICF), and other relevant study documents were submitted for review and approved by an appropriately constituted, Independent Central Institutional Review Board (IRB). Recruitment materials and site-specific modifications to the approved consent were reviewed and approved by the IRB before the site was initiated. eConsent was obtained from the participant prior to initiating the collection of relevant EHR data. As part of the eConsent process, participants were required to sign a statement of informed consent that meets the requirements of 45 CFR 46, local regulations, International Council on Harmonization (ICH) Good Clinical Practice (GCP) guidelines, Health Insurance Portability and Accountability Act (HIPAA) requirements, IRB and study site requirements, where applicable. The site staff countersigned the eConsent prior to initiation of study data collection. A participant was free to withdraw from the study at any time at his/her own request or could be withdrawn at any time at the discretion of the investigator.

As this trial was a pragmatic observational study with no IP and no required interventions, the investigators were not obligated to actively seek adverse event (AE) or serious adverse event (SAE) information. However, if the investigator learned of an SAE during the participant's participation in the study, the investigator or qualified designees were responsible for documenting, reporting, and following-up on the events according to their institutional policies and requirements.

4. Data collected

Variables were extracted for consented participants from the 3 different EHR systems at baseline and from their routine care encounters for 12 weeks.

Demographics: Age, Gender, Race, Ethnicity.

Vital Sign Measurements: Weight, Height, Blood Pressure.

Lab Results: HbA1c results.

Diagnoses: All relevant current diagnoses, new diagnoses, and changes during study period.

Medications: All current medications at baseline, new medications or changes during study period.

Clinical Encounters: Date and type of encounter with a healthcare provider.

Eligibility confirmation: Entered by site staff prior to eConsent.

eConsent: Participant provided electronic informed consent (eConsent) followed by site staff electronic countersignature.

ePRO: SF-12 at baseline and follow-up (12 weeks ± 2 weeks)

Following site eConsent countersignature, the participant was directed to complete the baseline ePRO survey. The follow-up ePRO was to be completed 12 ± 2 weeks. The variables were collected using the Optum Smart Measurement® System (SMS).

The study protocol stated 45 participants were expected for analysis. To account for a ~20% drop-out rate, we planned to enroll 57 participants. The end of study participation for each participant was 12 ± 2 weeks following the date of enrollment. Due to COVID-19 impacts on the sites, enrollment was concluded in October 2020 (at 48 participants), to allow follow-up activities to be completed in early 2021.

The following endpoints were acquired and evaluated at the conclusion of the study.

- Mean change in HbA1c from baseline to end of study
- Mean change in oral antidiabetic medications prescribed from baseline
- Mean change in participants' self-reported physical and mental health via the SF-12v2® Health Survey ePRO from baseline

All endpoints are summarized descriptively. Baseline and demographic characteristics are summarized by standard descriptive summaries [e.g., means and standard deviations (SD) for continuous variables or median and interquartile range (IQR) such as age and percentages for categorical variables such as gender]. The study was not formally powered; therefore, no statistical analysis was conducted.

5. Results

The study enrolled 48 participants from three sites [site 1: 15; site 2: 12 and site 3: 21]. As consent was obtained remotely, baseline data were extracted from the most recent historical encounters preceding consent, the metadata for those data recorded the date and time data were entered into the EHR.

Table 1
Schedule of events and data collected throughout study participation period.

Study Procedures	Baseline			Week 12 ± 2 week		
Informed consent (eConsent)	X					
Eligibility confirmation	X					
EHR data entry[a]	X			Daily through 12 weeks		
ePRO deployed (SF-12)	X			X		
RWD extracted (daily [b])	# subjects	# Records Collected [c]	# Data fields per record [d] (source)	# subjects	# Records Collected [c]	# Data fields per record (source)
Electronic eligibility confirmation	48	48	10	N/A	N/A	
Electronic informed consent	48	48	3	N/A	N/A	
Demographics	48	48	4	N/A	N/A	
Vital sign measurements	48	192	7	N/A	N/A	
Co-existing diseases [b]	48	841	3	20	48	3
Concomitant medications [b]	48	489	7	41	101	7
HbA1c[b]	48	94	6	33	33	6
Clinical encounters [b]	N/A	N/A	N/A	39	133	3
ePRO (SF-12)	48	48	N/A	28	28	N/A
TOTAL (2,151 Records)		**1,808**			**343**	
TOTAL (10,118 Data Fields)			**8,670**			**1,448**

EHR = electronic health record; ePRO = electronic patient-reported outcome; HbA1c = Glycated hemoglobin; NA= Not applicable.
[a] EHR data was collected daily as the encounters occurred and data was recorded by the clinical sites in accordance with the assessments performed during the routine clinical care visits.
[b] Data extracted daily as received from the EHR feeds.
[c] 1 Record = 1 row of data.
[d] Number of data fields that the site staff would have entered manually subject to source data validation.

The mean (SD) and median (IQR) age of the included participants were 58.7 (±10.8) years and 57.5 (50.8–65.3) years, respectively. Over 54% (n = 26) were female; 79% (n = 38) were White or Caucasian, 15% (n = 7) were African American; and 69% (n = 33) were non-Hispanic/Latino. Data was directly extracted from the entries within the EHR records. Although this study included only 3 sites, the diabetes cohort distribution in the entire network is comparable to the three sites participating in the study. Table 1 shows the volume of data extracted throughout the 12-week study.

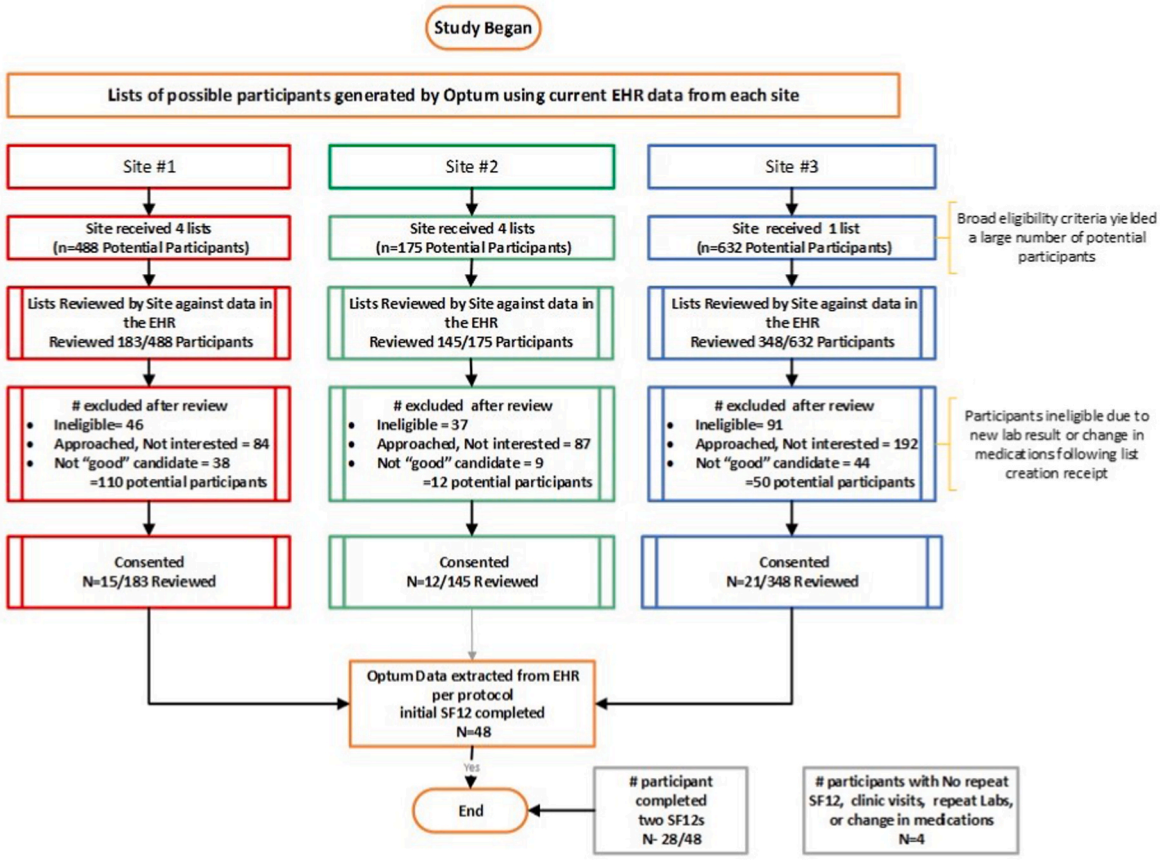

Fig. 5. Optum Proof of Concept Study Participation Consort diagram.

5.1. Screening efforts

Sites received lists of potentially eligible participants based on data available at the time of list generation. The site staff accessed their EHR to confirm eligibility, then approached the participant for enrollment in the study. The time expended for confirmation of eligibility was approximately 3 min per participant. Fifty-four percent (n = 363/676) of the eligible participants approached were not interested, primarily due to the lack of direct benefit (reported by site staff).

As there was a time lag between the date the participant lists were created and the date sites reviewed the list to confirm eligibility, 26% (n = 174/676) of potentially eligible participants had a medication change and/or a more recent HbA1c result which resulted in participant ineligibility. With this knowledge, Optum changed the frequency of generating subsequent lists to sites. An additional 13% (91/676) were determined by the site as 'not a good candidate' for reasons such as language barriers; cognitive impairment; or personal family reasons, etc. Refer to Fig. 5.

5.2. Medication changes

All enrolled participants were receiving metformin monotherapy at baseline, meeting the inclusion criteria. During the 12-week study period, 25% (n = 12) of participants had a change in diabetes-related medication; **8** had an additional anti-diabetic medication prescribed, **2** had their metformin formulation modified; **1** changed metformin dose; and **1** changed to a sitagliptin-metformin combination.

5.3. HbA1c changes

The most recent HbA1c result from 1 year prior to enrollment was collected as the baseline, and all subsequent HbA1c results during their 12-week follow-up were collected.

At baseline, the mean (SD) and median (IQR) HbA1c of the 48 participants were 7.82% (1.26) and 7.40% (7.20–7.83), respectively, and ranged from 6.9 to 12.7). During the 12-week participation period, 33 participants (67%) had ≥1 HbA1c test performed (there were no protocol requirements for follow-up). For these 33 participants' HbA1c, 21 decreased, 10 increased, and 2 were unchanged.

5.4. Clinical encounters

For the 48 participants, one hundred thirty-three (133) healthcare encounters were gathered from their EHR during the 12-week follow-up including 100 outpatient visits, 16 minor procedure/test, 12 telehealth/virtual visits, 3 urgent care, 1 hospitalization, and 1 major procedure/surgery.

5.5. QoL changes

The objective for inclusion of the SF-12 survey was to demonstrate the ability to merge an additional eSource to the database and load it to the EDC. No significant changes in SF-12 were expected or observed in the 28 participants who completed two SF-12's.

5.6. Data cleaning

All EHR data processed were subjected to multiple automated quality checks as the data moved through the system (Fig. 3), and each file was manually reviewed for quality prior to loading to EDC. For the entire study dataset of 10,118 data points were collected, and only 10 data queries were generated over the 8 months of data collection; 1 auto-generated failure due to field length incompatibility, and 9 manual queries were raised to the sites:

- 4 queries related to duplicate medication records
- 2 queries related to missing medication end dates
- 2 queries related to confirmation of HbA1c results
- 1 query related to a significant height measurement change between encounters

Five manual queries were addressed by an update to the EHR data, and 4 required a comment on the eCRF only as confirmation of data accuracy.

5.7. Final data review and verification by site staff

At the conclusion of the study the following processes were performed to verify the data extracted from the EHR were attributable to the source, legible, contemporaneous via EHR audit records, original, accurate and complete.

- The site staff reviewed the eCRFs via printed subject casebooks from the EDC system and reviewed the data in the EDC. This review identified no issues.

Optum verification of data included:

- Comparison of data in the EHR to data loaded to the Stage database
- Comparison of the data in Stage database to the Study database
- Comparison of the Study database to the EDC (via review of both the EDC UI PDFs by the sites, and data tables from the EDC system).
- Confirmation via MD5 hash algorithm of the match between the extracted and loaded files
- Review of all log files for errors during the transactions and transformations.

These comprehensive checks on the data confirmed that the data extracted for the study was accurate, complete, and traceable to the source data for all participants.

6. Discussion

This study was designed to test the ability to gather and extract all baseline and required ambispective study data via eSource EHR and an ePRO. We focused on a patient population who would receive clinically indicated interventions in a real-world care setting by targeting "not well controlled" T2DM participants being treated with metformin monotherapy. We targeted this population in anticipation that their routine care would result in healthcare visits, medication adjustments, and repeat laboratory tests to improve the control of their diabetes.

Study planning and system development began in November 2019, preceding the onset of the COVID-19 pandemic. When recruitment began in June 2020, a protocol amendment was approved to allow remote eConsent. During the 12-week follow-up, many routine healthcare visits were delayed due to COVID-19 restrictions, decreasing in-person interactions with healthcare providers, and resulting in lower-than-expected repeated HbA1c or medication adjustments. The study planned to enroll 57 participants but was stopped at 48 to allow for completion of the last participant's follow-up in early 2021. Enrolling participants into a study with no direct clinical benefit and performing outreach and consent remotely in the time of COVID lockdown was limiting. We believe that face-to-face interactions with the Principal Investigator or the clinical teams would have yielded a higher enrollment rate.

Approximately half of the participants did not complete their follow-up SF-12 survey. This may have been due to a lack of participant interaction and engagement with the site personnel (no scheduled visits, all study procedures were conducted remotely), discomfort with electronic procedures, lack of confidence in email reminders, lack of personal benefit from study participation, coincidental timing of the lifting of some COVID restrictions and return to work or calls from

unrecognized numbers due to the upcoming Presidential election.

The objective to perform an observational trial using the EHR to EDC mapping was successful. The only data entry required by the site staff was to confirm the participants' eligibility and eConsent. Participant identification for recruitment was eased by using the lists provided by Optum, significantly reducing manual EHR searches to find potential participants. Finally, verification of the eSource data at the end of the study (by both Optum and the sites) confirmed the data extracted from the EHR and loaded into the EDC was complete and accurate.

For any study destined to support a regulatory decision, sponsors have an obligation to perform source data verification (SDV) to ensure data in the EDC matches the data in the source. This process is time consuming and costly and can account for ~25% of the budget for a given trial [11]. Extracting the "source" data *and* the source meta-data directly from EHR and loading it to an EDC system reduces the need for SDV, eliminates the need for transcription and manual data entry, and significantly reduced the number of data queries generated.

6.1. Quantity and quality of EHR data extracted

For the 48 participants enrolled, a significant number of historical records of active medications and diagnoses were extracted (Table 1). Had this study required manual entry, the effort would have been significant, and site staff may have determined that some data was not important and omitted entry. We believe that the ease of pulling complete history via our methods allowed a more thorough view into a participant's history and could provide significant insights in some trial situations. Regarding the reduction of data queries, using a literature comparison, a study with 10,000 data points via manual data entry could have resulted in over 300 data queries [12], compared to the 10 queries generated for the study in question.

Our study resulted in significantly decreased work for all site staff and faster access to quality data in the EDC for a sponsor. All study procedures were conducted remotely including eligibility confirmation, eConsent and ePRO by the participants, available in near real time. Also, direct extraction of RWD occurred during regularly scheduled healthcare interactions supporting the goal of "research as a care option" and data copied from EHR through the Optum system to EDC was generally complete in 24–48 hours.

6.2. Limitations

The following limitations of the study are acknowledged; 1) the study did not offer translated consents, therefore, non-English speaking participants were excluded; 2) only 3 EHR systems were utilized for this study, therefore EHR systems not included may present differing data challenges; 3) medication adjustments that occurred within the 12-week follow-up did not permit subsequent HbA1c results to be obtained before the study end, therefore, corresponding changes in HbA1c were not available in the data; 4) sites with over several hundred participants on their initial list were hesitant to receive an updated list, therefore when confirming eligibility, participants were often found to not qualify due to newly available data (HbA1c too low or additional medication prescribed); 5) this unique process of loading all data directly to an EDC had never been done before and some challenges were identified that required correction mid-study.

7. Conclusions

Clinical trials are critical to the progress of medical science; however, management and execution of these trials pose significant challenges [13–15]. The task of identifying, screening and consenting participants for trials is labor-intensive. Advancements in eligibility identification are being developed [16–24], however, manual transcription of data from one system to another remains an obstacle. Clinical trials are the most expensive component of drug development, and improvements in their efficiency would be valuable [25]. Additionally, the expanding availability of digital health data and federal initiatives promoting the use of RWD [26] indicates this is the time to develop methods to extract EHR data for trials.

The overarching goal of this proof-of-concept study was to demonstrate that Optum could process RWD and verify EHR eSource data as fitting the requirements of Real-World Evidence (RWE) in a prospective observational clinical investigation. As defined in the Draft FDA guidance, a study must "demonstrate that each data source contains the detail and completeness needed to capture the study populations, exposures, key covariates, outcomes of interest, and other important parameters (e.g., timing of exposure, timing of outcome) that are relevant to the study question and design" [27]. Using EHR data from 3 disparate EHR systems from 3 geographically distant HCOs, Optum has demonstrated that this process could be scaled to more sites for future trials where EHR data are the appropriate source. Programmatically extracting EHR data eased the prescreening process by providing lists of potential participants, eased the site's data entry and query management burden, and made clinical trial data available to the sponsor within 24–48 hours of EHR entry. Gathering data and metadata from the source significantly reduced data clarifications and queries as well as the need for SDV.

The innovative processes deployed in this study have broad applicability for future clinical trials. Using EHR data to identify potentially eligible participants will ease site screening efforts and increase the ability of sites to report correct potential participant counts to study sponsors, allowing sponsors to select viable sites. Using EHR and other eSource data to populate eCRF fields can be performed for approximately 80% of the required clinical data (dependent on the relevant data to be collected), easing the site burden and query resolution effort typical after manual transcription. The clinical trial processes that have been in place for decades are ripe for disruption via the use of currently available technology. We believe that the solutions and processes the Optum DRN has developed can be successfully scaled to any number of sites and EHR systems within the network. Programmatically identifying potential participants and removing most data transcription requirements will have a significant impact on research performance as we move forward to implement them for more complex studies.

Authors' contributions

All authors contributed to the study conception, design, and execution. Cynthia Senerchia, David Wimmer, Samantha Cheng, Devin Tian, and Irene Margolin-Katz were responsible for the data extraction methodology, curation, and validation. Peter Payne and Cynthia Senerchia were responsible for the study supervision and Tracy Ohrt was responsible for the project administration. The investigation was performed by Brian Webster, Lawrence Garber, and Stephanie Abbott. All authors were involved in draft preparation and revision for important intellectual content. All authors approved the final version and agree to be accountable for all aspects of their work.

Funding

This research was funded by Optum [OptumInsight/Life Sciences] including the study design, conduct of the research, collection, analysis and interpretation of data, preparation of the article and the decision to submit the article for publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The following authors are direct employees of OptumInsight, the division of the company that funded the study described in this manuscript: *Cynthia Senerchia, Tracy L. Ohrt, Peter N. Payne, Samantha Cheng, David*

Wimmer, Irene Margolin-Katz, Devin Tian, The following author is an employee of Reliant Medical Group, a part of OptumCare, separate from the division of Optum (OptumInsight) that provided funding for the study described in this manuscript: *Lawrence Garber*. The following authors own stock or stock options in United Health Group, the parent company of the company that funded the study (OptumInsight): *Cynthia Senerchia, Tracy Ohrt, Peter Payne, Samantha Cheng*.

Acknowledgements

The authors would like to acknowledge the efforts and contributions of the members of the Digital Research Network team for their support of the project including Jignesh Patel, Mary Horan, and Subhashini Gopalan. We wish to acknowledge the research coordinators that contributed to the execution of the study including Robert Trent Holmes, Peggy Preusse, Patricia Arruda and Kimberly Nordstrom-McCaw. In addition, the DRN would like to acknowledge the efforts and commitments of the Quality Metrics team for deploying the eConsent and ePRO application and Adaptive Clinical for development and implementation of the unique electronic data capture system. Additional acknowledgements to Dr. Abhishek Tibrewal and Dr. Shailaja Daral for writing and performing the statistical analysis, and Esther Barlow for providing writing and editorial support.

References

[1] D.P. Byar, R.M. Simon, W.T. Friedewald, et al., Randomized clinical trials - perspectives on some recent ideas, N. Engl. J. Med. 295 (1976) 74–80, https://doi.org/10.1056/NEJM197607082950204.

[2] K.F. Schulz, D.A. Grimes, Generation of allocation sequences in randomised trials: chance, not choice, Lancet 359 (9305) (2002) 515–519, https://doi.org/10.1016/s0140-6736(02)07683-3.

[3] J. Concato, N. Shah, R.I. Horwitz, Randomized, controlled trials, observational studies, and the hierarchy of research designs, N. Engl. J. Med. 342 (25) (2000) 1887–1892, https://doi.org/10.1056/nejm200006223422507.

[4] V. Jenkins, V. Farewell, D. Farewell, et al., Drivers and barriers to patient participation in RCTs, Br. J. Cancer 108 (7) (2013) 1402–1407, https://doi.org/10.1038/bjc.2013.113.

[5] F.B. Fogel, Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review, Contemp Clin Trials Commun 11 (2018) 156–164, https://doi.org/10.1016/j.conctc.2018.08.001.

[6] P.M. Rothwell, External validity of randomised controlled trials: "to whom do the results of this trial apply?", Lancet 365 (9453) (2005) 82–93, https://doi.org/10.1016/s0140-6736(04)17670-8.

[7] J.R. Rogers, L. Junghwan, Z. Ziheng, et al., Contemporary use of real-world data for clinical trial conduct in the United States: a scoping review, J. Am. Med. Inf. Assoc. 28 (1) (2021) 144–154, https://doi.org/10.1093/jamia/ocaa224.

[8] Y. Zheng, S. Ley, B. Frank, Global aetiology and epidemiology of type 2 diabetes mellitus and its complications, Nat. Rev. Endocrinol. 14 (2) (2018) 88–98, https://doi.org/10.1038/nrendo.2017.151.

[9] M.E. Maruish (Ed.), User's Manual for the SF-12v2 Health Survey, third ed., QualityMetric Incorporated, Lincoln, RI, 2012.

[10] J. Ware, M. Kosinski, D. Turner-Bowker, et al., SF12v2: How to Score Version 2 of the SF-12 Health Survey, QualityMetric Incorporated, Lincoln, RI, 2002.

[11] S. Funning, A. Grahnén, K. Eriksson, Å. Kettis-Linblad, Quality assurance within the scope of Good Clinical Practice (GCP) – what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members, Qual. Assur. J. 12 (2009) 3–7, https://doi.org/10.1002/qaj.433. First published. (Accessed 26 February 2009).

[12] P. Stokman, et al., Risk-based quality management in CDM, Journal of the Society for Clinical Data Management 1 (1) (2020) 1–8, https://doi.org/10.47912/jscdm.20, 1.

[13] P.J. Embi, A. Jain, C.M. Harris, Physicians' perceptions of an electronic health record-based clinical trial alert approach to subject recruitment: a survey, BMC Med. Inf. Decis. Making 8 (2008) 13, https://doi.org/10.1186/1472-6947-8-13.

[14] S.R. Thadani, C. Weng, J.T. Bigger, et al., Electronic screening improves efficiency in clinical trial recruitment, J. Am. Med. Inf. Assoc. 16 (2009) 869–873, https://doi.org/10.1197/jamia.m3119.

[15] P.J. Emb, P.R. Payne, Clinical research informatics: challenges, opportunities and definition for an emerging domain, J. Am. Med. Inf. Assoc. 16 (2009) 316–327, https://doi.org/10.1197/jamia.m3005.

[16] A.J. Butte, D.A. Weinstein, I.S. Kohane, Enrolling patients into clinical trials faster using real time recruiting, Proc AMIA Symp (2000) 111–115.

[17] P.J. Embi, A. Jain, J. Clark, et al., Development of an electronic health record-based clinical trial alert system to enhance recruitment at the point of care, AMIA Annu Symp Proc (2005) 231–235.

[18] R.W. Grundmeier, M. Swietlik, et al., Research subject enrollment by primary care pediatricians using an electronic health record, AMIA Annu Symp Proc (2007) 289–293.

[19] F.L. Nkoy, D. Wolfe, J.W. Hales, et al., Enhancing an existing clinical information system to improve study recruitment and census gathering efficiency, AMIA Annu Symp Proc (2009) 476–480.

[20] S. Treweek, E. Pearson, N. Smith, et al., Desktop software to identify patients eligible for recruitment into a clinical trial: using SARMA to recruit to the ROAD feasibility trial, Inf. Prim. Care 18 (2010) 51–58, https://doi.org/10.14236/jhi.v18i1.753.

[21] S. Heinemann, S. Thüring, S. Wedeken, et al., A clinical trial alert tool to recruit large patient samples and assess selection bias in general practice research, BMC Med. Res. Methodol. 11 (2011) 16, https://doi.org/10.1186/1471-2288-11-16.

[22] T.R. Pressler, P.Y. Yen, J. Ding, et al., Computational challenges and human factors influencing the design and use of clinical research participant eligibility pre-screening tools, BMC Med. Inf. Decis. Making 12 (2012) 47, https://doi.org/10.1186/1472-6947-12-47.

[23] L. Penberthy, R. Brown, F. Puma, et al., Automated matching software for clinical trials eligibility: measuring efficiency and flexibility, Contemp. Clin. Trials 31 (2010) 207–217, https://doi.org/10.1016/j.cct.2010.03.005.

[24] C.C. Beauharnais, M.E. Larkin, A.H. Zai, et al., Efficacy and cost-effectiveness of an automated screening algorithm in an inpatient clinical trial, Clin. Trials 9 (2012) 198–203, https://doi.org/10.1177/1740774511434844.

[25] M. Dickson, J.P. Gagnon, Key factors in the rising cost of new drug discovery and development, Nat. Rev. Drug Discov. 5 (2004) 417–429, https://doi.org/10.1038/nrd1382.

[26] T. Hernandez-Boussard, K.L. Monda, et al., Real world evidence in cardiovascular medicine: ensuring data validity in electronic health record-based studies, J. Am. Med. Inf. Assoc. 26 (11) (2019) 1189–1194, https://doi.org/10.1093/jamia/ocz119. Erratum in: J Am Med Inform Assoc. 2020;27(1):181. PMID: 31414700; PMCID: PMC6798570.

[27] Us Food and Drug Administration, Real-world data: assessing electronic health records and medical claims data to support regulatory decision making for drug and biological products, Guidance for Industry (2021). Issued 0n September 30,, https://www.fda.gov/media/152503/download. (Accessed 12 March 2022).