Journal of Biomedical Informatics 117 (2021) 103765

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Creating and implementing a COVID-19 recruitment Data Mart

Tara T. Helmer [a], Adam A. Lewis [a], Mark McEver [a], Francesco Delacqua [a], Cindy L. Pastern [a], Nan Kennedy [a], Terri L. Edwards [a], Beverly O. Woodward [b], Paul A. Harris [a,c,*]

[a] Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA
[b] Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN, USA
[c] Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

ARTICLE INFO

Keywords:
COVID-19
Clinical Research
Research Recruitment
REDCap

ABSTRACT

The COVID-19 pandemic has resulted in an unprecedented strain on every aspect of the healthcare system, and clinical research is no exception. Researchers are working against the clock to ramp up research studies addressing every angle of COVID-19 – gaining a better understanding of person-to-person transmission, improving methods for diagnosis, and developing therapies to treat infection and vaccines to prevent it. The impact of the virus on research efforts is not limited to investigators and their teams. Potential participants also face unparalleled opportunities and requests to participate in research, which can result in a significant amount of participant fatigue. The Vanderbilt Institute for Clinical and Translational Research recognized early in the pandemic that a solution to assist researchers in the rapid identification of potential participants was critical, and thus developed the COVID-19 Recruitment Data Mart. This solution does not rest solely on technology; the addition of experienced project managers to support researchers and facilitate collaboration was essential. Since the platform and study support tools were launched on July 20, 2020, four studies have been onboarded and a total of 1693 potential participant matches have been shared. Each of these patients had agreed in advance to direct contact for COVID-19 research and had been matched to study-specific inclusion/exclusion criteria. Our innovative Data Mart system is scalable and looks promising as a generalizable solution for simultaneously recommending individuals from a pool of patients against a pool of time-sensitive trial opportunities.

1. Introduction

The COVID-19 pandemic has proven highly disruptive in the planning and implementation of biomedical research [1,2]. The unforeseen logistical difficulties that have plagued clinicians during this crisis, such as shifts in staff responsibilities, the need for physical distancing, and shortages of personal protective equipment [2,3], have also extended to researchers planning and conducting COVID-19 related trials. Furthermore, investigators studying COVID-19 have faced additional challenges unique to this area of research, due at least in part to the inherent urgency to quickly discover safe and effective measures to test, treat, and prevent the disease, often while attempting to balance and continue their pre-COVID-19 research priorities and responsibilities.

First identified in December 2019 [4], COVID-19 has been poorly characterized [5,6], lacking standardized codes for testing and diagnosis prior to March 2020 [7]. Initial research favored single-arm or observational studies over randomized clinical trials [8,9], and the rush to publish results led to questionable ethical choices in research methods [8,10] and in some cases retracted manuscripts [11]. As research ramped up, a flood of trials began entering the pipeline with initial funding from institutional sources or grant supplements, then giving way to larger dedicated coordinated trials. Over the next several months, the National Institutes of Health (NIH) responded to the crisis by re-appropriating at least two well-established clinical trials networks within the National Institute of Allergy and Infectious Diseases (NIAID): The HIV Vaccine Trials Network (HVTN) and the AIDS Clinical Trials Group (ACTG) evolved into a new and, it is hoped, temporary COVID-19 Trials Network into which complex trials of investigational vaccines and treatments could be rolled out quickly across multiple existing clinical research sites. Academic medical centers were faced with unprecedented competing demands on their research ecosystems. Healthcare IT teams, already overtaxed to support increased COVID-19 clinical informatics needs, were called in to assist with pressing requests from the research enterprise [12]. In addition, data emerging early in the

* Corresponding author at: Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA.
E-mail address: Paul.a.harris@vumc.org (P.A. Harris).

https://doi.org/10.1016/j.jbi.2021.103765
Received 15 October 2020; Received in revised form 5 March 2021; Accepted 27 March 2021
Available online 30 March 2021

pandemic on racial and ethnic disparities in COVID-19 infection rates and outcomes [13] underscored a critical need to find ways to support enrollment of underrepresented minority participants in COVID-19 clinical trials [14].

The observed growth of COVID-19 submissions to ClinicalTrials.gov and to our own internal IRB system suggested to our team at Vanderbilt University Medical Center (VUMC) that we would ultimately face the problem of multiple studies simultaneously recruiting from the same participant pool [15]. This presented additional complexities, as research teams had historically recruited and worked in a somewhat isolated fashion without the need to co-coordinate or 'share' a population of eligible research patients. Initially, our process at VUMC involved having attending physicians form 'morning huddles' to guide inpatient prioritization for consideration and assignment to specific recruiting studies. Outpatients, however, not being centrally located, posed a new problem–how to optimize participant-to-trial matching for a population not amenable to our internal workflow. Furthermore, the need to identify potential research participants within hours of their COVID-19 diagnosis presented a unique challenge to researchers testing outpatient treatments and highlighted a critical need to partner with biomedical informatics specialists in order to leverage all available systems to fight the pandemic. We needed to formulate a more efficient system that would ensure outpatients also had an opportunity to participate in important, targeted COVID-19 research projects.

As an academic medical institution, VUMC has long embraced a goal to ensure that every patient has an opportunity to participate in clinical research. Coupling this aspiration with the desire to support as many competing COVID-19 trials as possible, we compiled a team from the VUMC research enterprise that included informaticians, research support experts, compliance professionals, and research operations key informants. Together, we designed and built a Data Mart for COVID-19 research recruitment with the goal of identifying the 'right patient at the right time for the right COVID-19 study'. This Data Mart platform leverages REDCap **(R**esearch **E**lectronic **D**ata **Cap**ture) [16,17] and is supported by a combination of data feeds (including Fast Healthcare Interoperability Resources [FHIR][18] endpoints) from our Epic electronic health record (EHR) system. The Data Mart application aggregates this information along with patient 'consent to contact' (C2C) permissions collected from operational Data Mart workflow, then automatically generates participant 'goodness of fit' assessments for a pool of actively recruiting trials. The Data Mart also utilizes a 'virtual huddle' model to support the sharing of potential matches across studies. While originally formulated to allocate participants to outpatient trials, our method is suitable for inpatient trials as well. Here we describe our Data Mart platform and early results of implementation, along with a set of generalizable 'lessons learned' that may prove beneficial to other teams faced with similar recruitment and logistical challenges.

2. Material and methods

The design, development, and deployment of a recruitment-oriented Data Mart platform required input and support from VUMC policy teams, health equity experts, Human Subjects Protection experts, and Privacy Office leadership. Overall Data Mart leadership, technical programming, project management, and research stakeholder engagement efforts were provided by the Vanderbilt Institute for Clinical and Translational Research (VICTR). The platform was developed in an iterative fashion, evolving over time to support needs articulated by the research community most engaged with ongoing or projected COVID-19 trials. Details about the technical platform, data sourcing, and end-user process design details are provided below.

2.1. Defining Data Mart system requirements

In designing the Data Mart platform, we considered three essential goals: 1) evaluate every patient testing positive for COVID-19 for consideration as a candidate for one or more trials; 2) democratize access to the Data Mart matching service for as many research studies as possible; and 3) respect VUMC's 'opt-in' policy, which allows patients to choose whether or not to be contacted directly for research studies outside of normal clinical provider encounters.

2.2. Choosing a technical platform

Given resource constraints and a design requirement to ensure the system could rapidly evolve over time, we chose REDCap [16] as our central platform for data aggregation, management, and reporting workflow. REDCap is a secure web application developed and disseminated by our VUMC team and is available at no cost to academic, nonprofit, and government organizations around the world [17]. Whenever possible, we used native functionality within the REDCap system to support Data Mart data collection and workflow. In cases where Data Mart data integration or complex workflow requirements were not feasible with standard REDCap functionality, we developed REDCap External Modules. External Modules are add-on software programs that extend REDCap's current functionality or allow customizations and enhancements for REDCap's existing behavior and appearance. Software developers use GitHub for hosting, with externalization to the global REDCap community through a common dissemination repository [19].

2.3. Data sourcing

For clinical data sourcing, we prioritized use of Epic HL7-FHIR endpoints whenever available, supplementing with information from a centralized VUMC Health Data Repository (HDR) [20] and Research Data Warehouse (RDW) [21] as needed. The HDR was especially useful at the beginning of the pandemic when COVID-19 lab results were not readily available from Epic. Later, we adjusted our pipeline to retrieve results directly from Epic, but we still use the HDR as a means of identifying tested patients before results are completed and available.

2.4. High-level Data Mart system architecture

Fig. 1 provides a high-level overview of our system architecture. Clinical and research participation data are derived from several Epic-based (FHIR and non-FHIR) resources, while real-time event and data notifications trigger from the VUMC HDR. Derived and extracted data necessary for Data Mart operation are stored in two REDCap projects. A REDCap External Module supports the presentation layer for VICTR staff and research end-user support.

2.5. Capturing participant preference for contact

A central feature of the Data Mart model is *a priori* capture and recording of individuals who are interested and willing to receive direct contact from researchers about new COVID-19 research opportunities. VUMC is an 'opt-in' research contact institution, requiring special permission from patients to be contacted by individual research teams if contact is made outside normal clinical encounters.

Evaluation of numerous study protocol inclusion requirements for time-to-enrollment after COVID-19 testing and diagnosis highlighted a need to expedite the opt-in mechanism. Many outpatient treatment trials would require patients to begin treatment within days of testing positive for COVID-19, so the system needed to support a workflow in which patients had the opportunity to consent to contact *before* their diagnosis was confirmed. Starting a "Consent to Contact' (C2C) permission request after diagnosis would require too much calendar time, thereby eliminating a high number of patients from consideration for COVID-19 trials. Given this, we decided to initiate the process as upstream as possible by sending the C2C message to patients immediately after they were tested

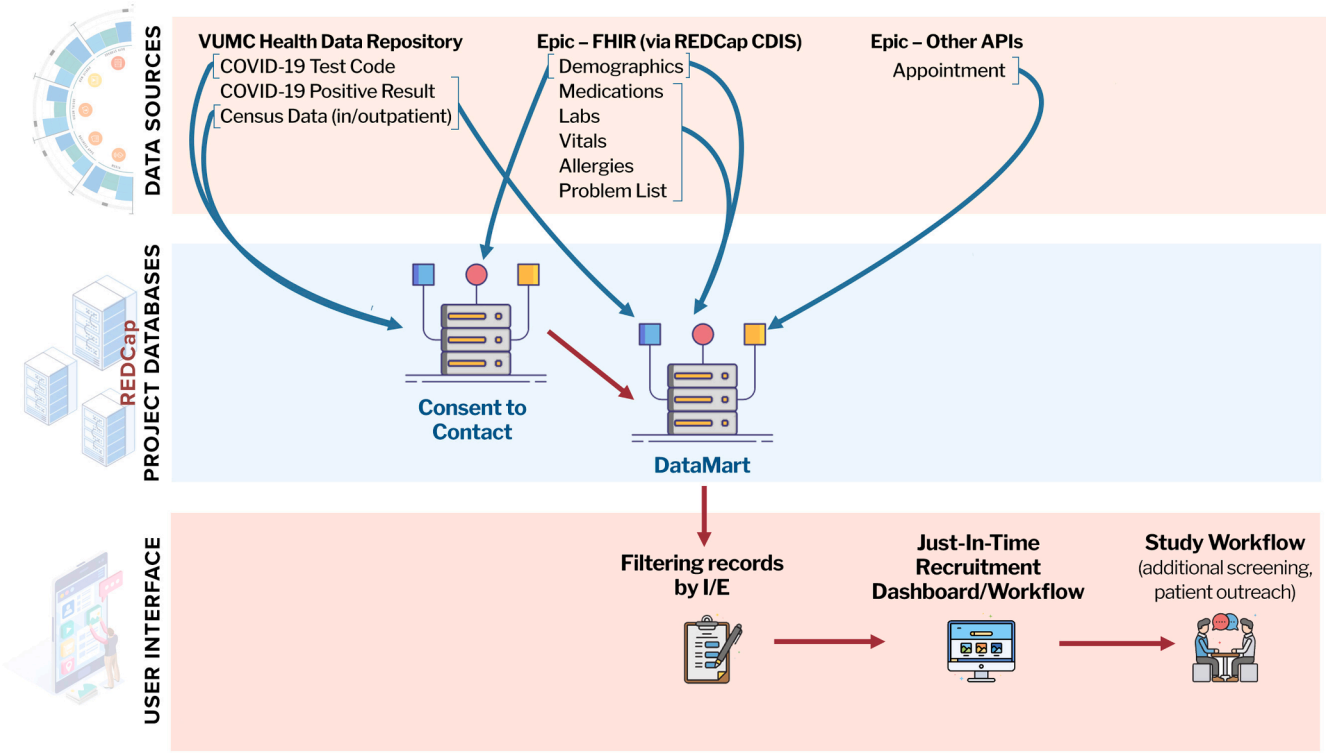


Fig. 1. The system architecture highlights the interoperability across various systems and sources.

(with event triggered by clinical system notification of a COVID-19 test), rather than waiting for confirmed results from laboratory systems. Shifting the time for messaging allows patients not only to share their preference for contact about COVID-19 research, but it also introduces them to the idea of COVID-19 research earlier. Finally, it also allows researchers to contact them within hours of the confirmed positive test result.

To facilitate the C2C workflow, we used an application programming interface (API) data feed from the VUMC HDR to detect when a new test was recorded within the EHR. Upon receipt, a medical record number (MRN) is inserted in a special 'C2C' REDCap project designed to immediately and automatically pull basic demographic contact information (patient name, e-mail address, contact phone number, and preferred language) via HL7 FHIR endpoints. REDCap then prepares and sends to all outpatients an automated message from Vanderbilt's Executive Vice President for Research. The default C2C message is created as an email, in English or Spanish depending on the patient's known preferred language as collected in the EHR. If an email address is not available in the EHR but a mobile number exists, the C2C database sends an SMS message. Every message includes contextual information and a link to a one-question survey, available in English or Spanish, asking the patient if they would like to be contacted directly for COVID-19 research opportunities at Vanderbilt. Responses are recorded in the C2C database. See Appendix A for examples of English version email and SMS message.

2.6. Capturing and recording COVID-19 positive test results

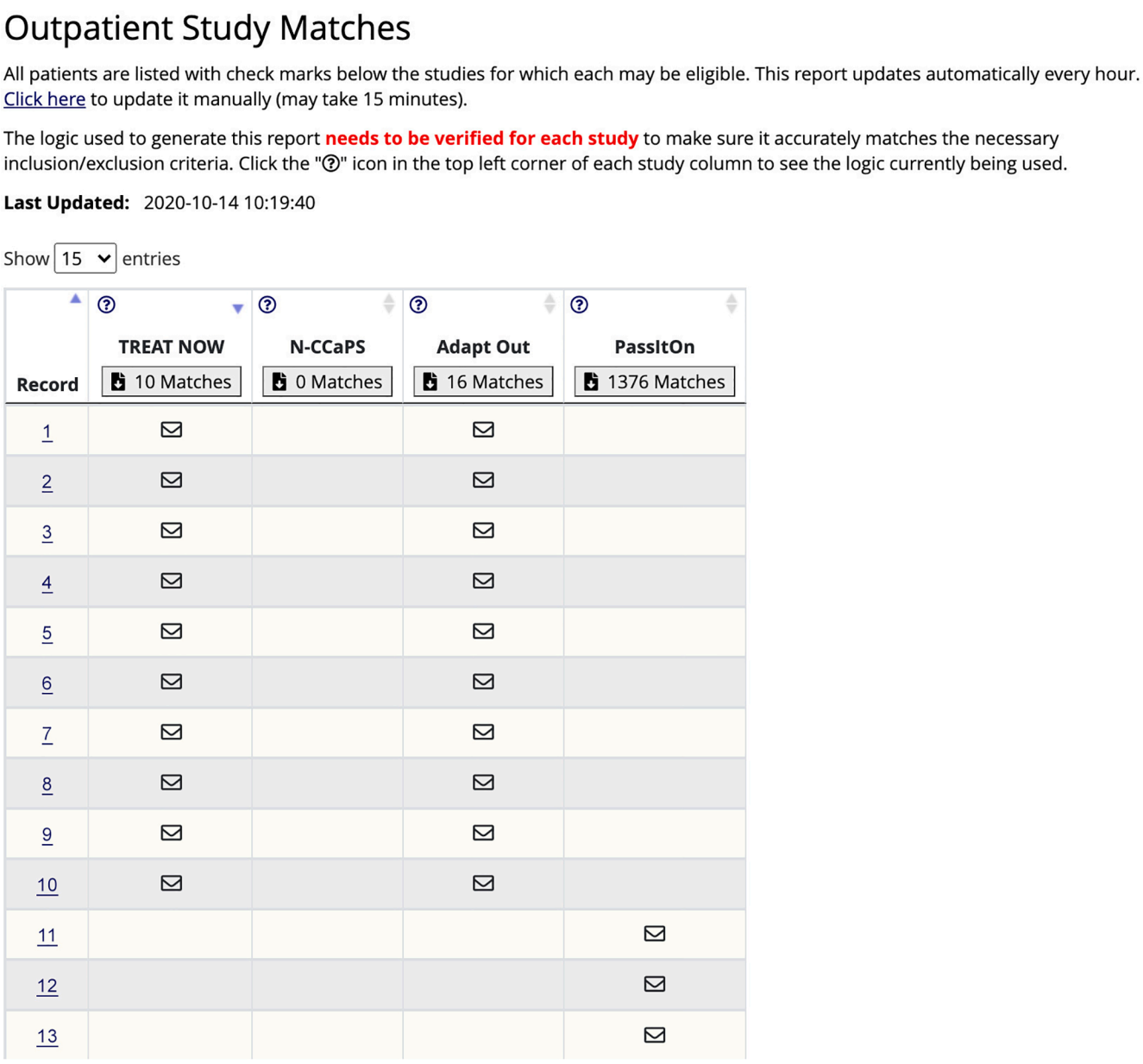
To complement the C2C project, REDCap 'listens' to another VUMC HDR data feed for real-time detection of any new positive lab test result for COVID-19. When detected, the patient's MRN is registered in a second database (Data Mart). FHIR and non-FHIR data sources are then used to automatically extract demographics, laboratory results, vitals, medications, allergies, problem list, and inpatient/outpatient status - from the EHR for immediate deposit in the Data Mart. Specifically, data are extracted that are relevant to known inclusion and exclusion criteria for Data Mart supported projects. Finally, the Data Mart program automatically refreshes each known COVID-19 positive patient with an 'OK to Contact' flag, collecting this information from the aforementioned C2C database. This repeating check ensures individuals who test positive before consenting to contact are included as soon as they consent, and also allows a mechanism to reset a patient's status if they change their mind later (either way) about being contacted.

2.7. Creating a study eligibility dashboard

For each participating trial, we code key inclusion/exclusion criteria rules to run against EHR-derived data for every potential participant who has agreed to direct contact. These criteria are informed by the study protocol with input from the PI/study team and driven by what is computable using retrievable EHR data. The computable phenotype coding work is performed in a REDCap External Module using a modular approach, which allows rapid deployment or sunsetting of individual COVID-19 studies. VICTR Research Support Services (RSS) team members use a dashboard in the REDCap system to view 'just-in-time' recommendations for suitability of each patient eligible to participate in each supported trial. The dashboard snapshot view quickly and easily provides the following information: study short name, the total number of current potential matches for that study, which patients have already been shared with the study team (indicated by an envelope icon), which patients have not yet been shared (indicated by a check mark), and a '?' icon that links out to view study specific logic. (See Appendix B for an example of study-specific logic.) Fig. 2 provides a partially redacted view of the Dashboard where eligible opt-in COVID-19 positive patients are listed as rows and each recruiting study opportunity is listed as a column.

2.8. Onboarding and support process for each new project

The process for onboarding a new project requires multiple steps (Fig. 3) and is initiated with a REDCap intake form to capture pertinent study information. Upon receipt, the request is reviewed for



Outpatient Study Matches

All patients are listed with check marks below the studies for which each may be eligible. This report updates automatically every hour. Click here to update it manually (may take 15 minutes).

The logic used to generate this report **needs to be verified for each study** to make sure it accurately matches the necessary inclusion/exclusion criteria. Click the "?" icon in the top left corner of each study column to see the logic currently being used.

Last Updated: 2020-10-14 10:19:40

Show 15 entries

Record	TREAT NOW	N-CCaPS	Adapt Out	PassItOn
	10 Matches	0 Matches	16 Matches	1376 Matches
1	✉		✉	
2	✉		✉	
3	✉		✉	
4	✉		✉	
5	✉		✉	
6	✉		✉	
7	✉		✉	
8	✉		✉	
9	✉		✉	
10	✉		✉	
11				✉
12				✉
13				✉

Fig. 2. Snapshot of the Data Mart Dashboard provides key information at a glance.

completeness and appropriateness. VICTR RSS reviews the study-specific inclusion/exclusion criteria to determine compatibility with available data fields in the Data Mart. During an iterative process, VICTR RSS works with the PI/study team to finalize the criteria to drive the Data Mart logic. The goal is to provide the study teams with a report that gets them close to their target population, with the understanding that additional screening in the EHR will be necessary before a potential participant is deemed eligible and contacted.

In addition to the study-specific eligibility criteria, additional standard logic is included for every study, for example, to include patients who agreed to direct contact and to exclude all inpatients. With the logic finalized, the study specific external module is added to the Data Mart and appropriate IRB approval is confirmed. We also developed standardized template language to describe the Data Mart, which research teams can use for IRB submissions.

For any competing trials, the respective PIs/study teams are required to meet with VICTR RSS in a 'virtual huddle' to discuss equitable distribution of potential matches across studies before potential matches are shared. In that huddle a schedule is developed for sharing patients equitably across each of the competing trials. Currently a 2-day rotating schedule is in place. Each day a single study is designated as primary and gets right of first refusal; the other study is designated as secondary. The list of potential matches is sent to the primary study in the early morning. The primary study is expected to screen these patients and indicate their enrollment status in Epic – for example, Identified, Interested, Ineligible, Declined, On Study. The list of potential matches, including any overlapping patients, is then sent to the secondary study around mid-day. The secondary study team will initially check for any existing enrollment statuses in eStar. If a patient is designated as Identified, Interested or On Study they will not continue to screen nor will they try to contact the patient. This is critical to minimize participant fatigue and participant confusion. If a secondary study checks that patient later and sees an enrollment status of Ineligible or Declined, they can proceed with additional screening and if applicable, contact the patient. The use of enrollment statuses is critical in promoting transparency across studies and minimizing participant fatigue.

Once the study has been added to the Data Mart and the expectations have been outlined and agreed to by the study teams, the 2-step process

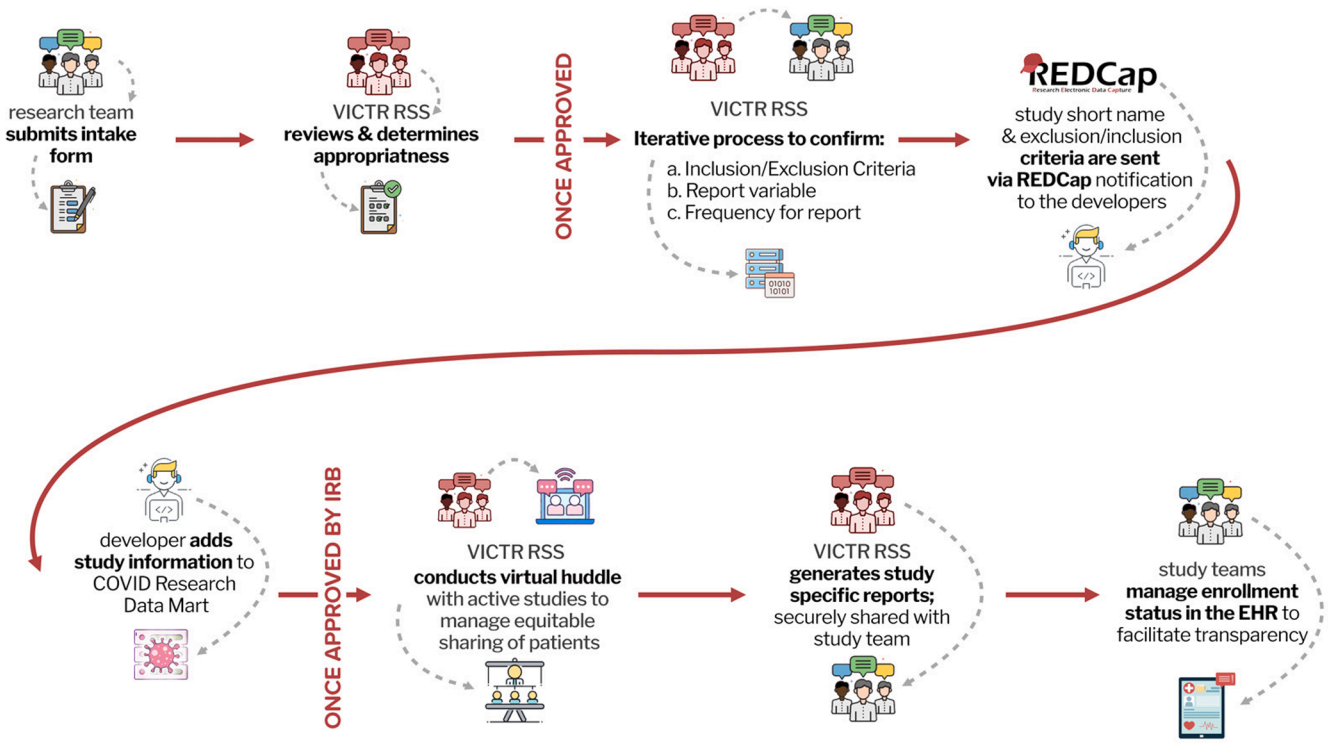


Fig. 3. Administrative support is crucial to facilitate operational workflow.

for identifying and sharing potential matches can begin. VICTR RSS reviews the dashboard for each record flagged with a check icon and marks each of those individual records to indicate the patient has been identified as a potential match for the respective study. This results in the check icon being updated to an envelope to indicate the patient has been sent to the study team. Next, VICTR RSS clicks on a button within the Outpatient Study Dashboard to export an excel file containing all the current matches for the respective study. The file, which contains patient MRN, name, date of birth, email address, phone number and date of last COVID positive test, is sent to the study team in a secure manner. For all subsequent reports, one additional step involves VICTR RSS removing any patients marked as 'Sent to Study' from the exported file before sharing it with the study team. In total, VICTR RSS spends about 20 min a day reviewing the dashboard, generating reports, and sharing the reports with the study teams.

After the daily report of potential matches is sent, the workflow shifts from VICTR RSS and the Data Mart platform to the study team. Once the study team receives its list of potential matches, they conduct additional screening via an electronic chart review in Epic to confirm study eligibility. One key aspect of their chart review is to confirm if the patient has any known current or recent research contact or enrollment activity. This review is crucial to help minimize participant fatigue. For example, if a research team is screening for an interventional trial and they confirm the patient is already on an interventional trial, they would not proceed as that would be a known reason for exclusion.

3. Results

3.1. Consent to contact model

We launched participant C2C operations on May 1, 2020. In the first 5 months (5/1/2020–10/9/2020), testing awareness generated automated messages to 62,427 individuals who were tested at the main hospital or one of 22 VUMC Testing Centers. The vast majority of these messages were English version emails (see Table 1).

Table 1
Breakdown of C2C messages sent by type, response, and response rate.

Message format	Would you like to be contacted directly about potential research projects related to COVID-19?		
	Yes	No	No response
English email (43,287)	5,537 (13%)	3,450 (8%)	34,300 (79%)
Spanish email (408)	19 (5%)	11 (3%)	378 (92%)
English SMS (7,059)	260 (4%)	248 (3.8%)	6,551 (92.2%)
Spanish SMS (368)	20 (5%)	14 (4%)	334 (91%)

3.2. Quantitative feedback - Data Mart

We launched the study-specific workflow, external module, and sharing of potential matches, on July 20, 2020 for a single study (TREAT NOW), and have since launched three additional studies:

- *Trial of Early Antiviral Therapies during Non-hospitalized Outpatient Window (TREAT NOW) for COVID-19* (NCT04372628) is an outpatient interventional trial for COVID-19 positive patients. In addition to logic involving medications, allergies and diagnoses on the Epic Problem List, a key requirement for identifying potential matches is a time window of less than 4 days from the date of their COVID test.
- The second study to launch was *COVID-19 in Cancer Patients Study (N-CCaPS): A Longitudinal Natural History Study* (NCT04387656). The main criteria used to identify potential matches, in this case patients undergoing active cancer treatment, is a scheduled appointment in an oncology clinic within the next 90 days.
- The third study to launch was *ACTIV-2/A5401: Adaptive Platform Treatment Trial for Outpatients with COVID-19 (Adapt Out COVID)* (NCT04518410), another outpatient interventional trial. Like TREAT NOW, the logic included a tight time window of less than 7 days from the date of their COVID test. This study also represented our first competing trial as most of the potential eligible patients overlap with the TREAT NOW study.

- The fourth study to launch was *Collection of biological samples of Convalescent patients from COVID-19*. This study supports the Passive Immunity Trial Of the Nation for COVID-19 (PassItOnII) through the recruitment of recovered participants to donate convalescent plasma. This study was a great complement to the outpatient interventional trials as it is focused on participants > 28 days since their positive test result.

Of note, ACTIV-2/A5401 and PassItOnII have been deemed high priority studies for Operation Warp Speed, which is a partnership among several federal government agencies to accelerate the development, manufacturing, and distribution of COVID-19 vaccines, therapeutics, and diagnostics, including the specific goal of delivering 300 million doses of a safe, effective vaccine for COVID-19 by January 2021 [22].

We report in Table 2 early results for generating study specific potential matches and limited enrollment outcome metrics from each of these four studies. There are many confounding factors that determine final eligibility and willingness to participate and enroll in specific studies. An examination of these factors is beyond the scope of this report, so we are reporting here only high-level enrollment statuses based on researcher reporting on their study-specific screening logs. We report two columns for enrollment (Data Mart attribution + total enrollment) to illustrate that the Data Mart is complementary with other methods of recruitment. Ongoing studies are also pursuing enrollment via clinician referrals, participant-facing printed recruitment materials, and other methods.

3.3. Qualitative feedback - Data Mart

We asked research teams to provide qualitative feedback on the utility of the Data Mart, recognizing that the positive utility might be measured in at least two ways: 1) increasing the rate of recruitment; and/or 2) reducing the amount of work needed to support recruitment. Qualitative feedback provided from research teams focused mainly on time saved, both in identification and screening of potential participants. The C2C workflow saves significant time by eliminating the need to go through a patient's healthcare provider to approach them about a potential research opportunity. Additionally, the algorithm-based narrowing of the COVID-19 population by study-specific eligibility criteria saves research coordinator time by reducing the volume of needed chart reviews.

One research nurse summed it up nicely, "The Data Mart has saved us hours in this process by filtering for many of our exclusions on the front end and highlighting the patients who are more willing to participate in research." Another stated, "The fact that we have 6 individuals on our radar who are eligible at this very moment is significant (aka amazing) – there is no way we would have been able to identify these 6 people quickly enough had we been tasked with reviewing all the charts of people who tested positive."

Table 2
Quantitative Information about initial four studies.

Study Name	Date Launched in Data Mart	Total Number of Potential Matches Sent	Number of Data Mart Potential Matches Enrolled	Total Number Enrolled During Same Time Frame
TREAT NOW	7/20/2020	279	10	28
N-CCaPS	8/27/2020	3	0	0
ACTIV-2/ A5401 (Adapt Out COVID)	9/5/2020	108	2	5
PassItOnII	9/21/2020	1300	50	94

3.4. Operational scalability of onboarding new studies

In addition to the study-specific results, the ability to provide equitable access combined with rapid deployment of new studies from both administrative and technical perspectives was crucial to our evaluation of the Data Mart. The administrative process – from study intake, to review, to finalizing eligibility criteria and communicating those to the software developers – has been streamlined using REDCap. Similarly, data and technical tools necessary to implement basic criteria have been developed and are available to support the implementation of each new study. The main variant across studies is the time it takes to translate inclusion/exclusion criteria into strict logic based on available data. While this can vary significantly based on criteria complexity, it has averaged about an hour for the last couple of studies. Once logic is finalized, the time required to add a new study is approximately 15 minutes. Provided all regulatory approvals are in place, new studies can now be onboarded and implemented in as quickly as 1–2 business days.

4. Discussion

The COVID-19 Data Mart is operational and has been successfully deployed to support screening and recruitment efforts for four concurrent trials. The Data Mart promotes equitable distribution of potential participants among researchers, alleviating competition. Importantly, this also prevents or at least minimizes the extent to which patients are contacted simultaneously by different teams about different studies, which could result in a confusing or frustrating patient experience. This is only possible if study investigators are willing to collaborate and cooperate, which is an expectation for studies wishing to use the Data Mart. In exchange for their collaboration, researchers gain access to a pool of patients who have already been screened on key eligibility criteria, thus increasing recruitment speed and efficiency. While this is important for every trial, the ability to provide an additional tool to support rapid participant identification is especially critical for high priority studies, such as those supported by Operation Warp Speed.

The flexible REDCap infrastructure and External Module framework has been instrumental in building and refining our Data Mart model. It allows fast, iterative development, and the ability to pivot quickly to support new ideas. For example, initially we messaged only patients who tested positive for COVID-19. After several weeks, however, we realized the need to connect with patients further upstream in the process to minimize recruitment delays for the time-sensitive studies, and so began messaging immediately after a COVID-19 test was documented–without waiting for the result. The REDCap structure has also enabled our technical methods to be scalable. Inclusion and exclusion criteria for each COVID-19 trial are encapsulated in an External Module, so that new studies can easily be onboarded–or completed studies removed–without potentially compromising the existing application architecture supporting other enrolling trials. The process of coding study-specific inclusion and exclusion criteria starts with assessment by an RSS Project Manager with input from the study team. Once specifications are defined, building the computable phenotype and activating the study in the Data Mart requires approximately ten minutes of programmer time. While we have considered building a self-service researcher-oriented interactive query interface, we are not convinced the added efficiency gain would be worth the additional investment in programming and training resources for this project.

The Data Mart fosters a partnership between VICTR RSS and each COVID-19 study team. Serendipitously, we have seen that the model also promotes coordination and collaboration among the study teams. Continual, joint monitoring and communication ensures that the right patients are being identified by the pre-defined criteria, or whether adjustments are needed. Once competing studies have been onboarded, study teams meet in a 'virtual huddle' with VICTR RSS to agree on an equitable allocation of potential participants. By working together on the front-end, research teams have learned about one another's trials,

which has allowed them to refer patients to one another and to introduce other research options or differentiate their study when discussing research with patients.

Transparency is a critical component of our process and is primarily facilitated by the tracking of patient enrollment status in Epic. A participant who enrolls in an interventional trial is marked accordingly in Epic, and is not invited to join another interventional trial, thus helping to mitigate participant fatigue. Similarly, tracking enrollment status in Epic allows research teams to determine whether the patient has previously been contacted or declined consent for another study, which are important details to acknowledge when contacting the patient about a subsequent research opportunity.

While facilitating screening and recruitment for COVID research was our primary goal, developing a system to support inclusion among underrepresented minorities was essential. We designed our process so that the original C2C message would be delivered in either English or Spanish, depending on the preferred language noted in a patient's EHR. Also, noting that Blacks and Hispanics are roughly twice as likely as whites (23% and 25% vs. 12%) to rely on smartphones for internet access [23], and are more likely than whites to use a smartphone to access health information [24], we built our C2C project to support SMS messaging in addition to email. We are currently exploring additional languages for the C2C invitation and estimate we will be supporting Arabic in the near future.

Although originally designed to help researchers, our Data Mart also benefits patients, particularly those who may be emotionally overwhelmed at the time of diagnosis or too sick to make a confident choice among multiple studies. An invitation to participate in a COVID-19 clinical trial, delivered soon after a positive test, may provide a message of hope to the patient and family in a difficult time. The C2C message is likely the first trial-related communication received by potential participants, and thus we strived to make it welcoming, informative, clear, and action oriented. We have revised it since the initial launch based on feedback from patients, our Privacy Office, and our Patient Education Department.

We are hearing from research teams that computable phenotypes and patient-trial matching algorithms are saving time for research coordinators over traditional chart review processes. This indirectly supports patient privacy because the computable matching process reduces the number of chart reviews that would otherwise be needed to filter eligible patients.

The Data Mart has proven successful and valuable for recruitment of COVID-19 trial participants, but we recognize there is no single perfect solution to identification of participants for all clinical trials. We have several other systems in place at VUMC, including native Epic research tools, ResearchMatch [25], TrialsToday [26], and SubjectLocator [21]. Each system adds collective value to our portfolio of researcher-enabling tools, and all are complementary. In the case of COVID-19, the new Data Mart model allowed us to solve two unique challenges: triggering preparatory C2C between lab test order and lab test result; and a novel method of simultaneously considering and recommending individual patients for multiple concurrently recruiting trials.

4.1. Limitations

While trying to make our system as inclusive as possible, we excluded two populations in favor of rapid development and implementation. Privacy safeguards at VUMC render the opening of employee EHR records more complex and time-consuming than for external patients. For the sake of speed, we decided to forego sending our C2C message to employees for now. While we initially included messaging for our pediatric population, the absence of pediatric-focused studies requesting to use the Data Mart resulted in an early pause on outreach to this population. Technically, moving to include both employees and pediatric populations will be straightforward, so the Data Mart platform is ready if and when policy decisions are made to support inclusion.

While the percentage of non-responders to our C2C message is fairly high, we do not know how many of the messages are undeliverable or are being filtered out by email servers. The large majority of COVID-19 tests at VUMC have a negative result (88%); many people are being tested who have no symptoms of COVID-19 and no reason to believe they have been infected. These patients may view the C2C message as irrelevant. Also, we have not yet translated the C2C message into languages other than English and Spanish; thus, we are missing some potential participants who speak other languages.

Despite positive feedback, we recognize the Data Mart is not the lone or ultimate solution for participant evaluation and recruitment into COVID trials. The ACTIV-2 study team has relayed that the 'declined to enroll' rate for complex interventional trials like ACTIV-2 is consistent between the Data Mart and other recruitment methods. Nevertheless, the ACTIV-2 study team also expressed that the process has saved immeasurable time by filtering out ineligible participants and by initiating the C2C process. With incredible limits and pressure on research staff and workloads, the Data Mart has eliminated much of the work required on the front-end of patient recruitment by providing teams with a cohort of likely eligible patients who have already expressed a willingness to be contacted about research.

As a result of Vanderbilt's research contact opt-in policy, we are not using all of Epic's native functionality. We recognize that these tools might partially negate the need for select components of our model, especially the C2C invitation workflow. That said, public interest in COVID-19 research is high [25,26], so our rates of interest in C2C where we are proactively triggering invitations on an event (testing) may be higher than we would see in a more general request framework. Our computable approach to matching a pool of patients to a pool of potential trials also ensures a high level of flexibility and customization with minimal IT investment on a per-study basis.

Our current participant eligibility distribution workflow is dependent on RSS personnel to serve as the honest broker and share daily matching reports with each respective study. This requires someone to log into the Data Mart, mark each new 'match' for the respective study, export the matching report and share it with the respective study team in a secure manner. While the level of effort at this point is minimal (20 min per day in serving four concurrent trials), eliminating this 'honest broker' dependency would make the model even more scalable. We consciously made the decision to implement our facilitated honest broker model over a purely self-service researcher workflow for expediency. We are learning lessons in our current work with COVID-19 research teams that will be helpful in making design decisions should we decide to later build a fully automated system.

4.2. Lessons learned

The concepts and lessons learned from developing and implementing our Data Mart platform should be transferrable and valuable to others designing or considering adoption of similar systems.

(1) Transparency, trust, and teamwork are foundational requirements for leading-edge innovation projects like this one. Multiple concept discussions were needed with regulatory and compliance leaders before considering launch. Our culture at VUMC favors innovation and is based on many years of cross-team collaboration. Given this high level of trust, institutional leaders are willing to support multiple iterations of concept clearance when needed. The patient messaging for our 'consent to contact' language is a great example and has gone through several iterations and review cycles from the Privacy Office and Patient Education Department.
(2) Technical architecture and tools are important, but they comprise only a part of the process and its success. Administrative support is crucial, as is the willingness of study teams to collaborate and

be as transparent as possible about their outreach to potential matches and the subsequent outcome.

(3) In an evolving pandemic, perfect is the enemy of good and there is no way to confidently define long-term system requirements and specifications. We focused all energy on solving a perceived (and later confirmed) need, then allowed our methods to evolve over time. In some cases, a short-cycle sprint of work highlighted errors in assumptions or unexpected dependencies, so we were forced to backtrack after gaining more information from one or multiple stakeholders. In all cases, we considered the platform to be an evolving solution and our choice of REDCap as an architectural platform allowed very rapid technical prototypes that were easy to evolve or rebuild.

(4) Lower-tech solutions are often very useful when timing is critical, and assumptions are untested. An example of how this strategy worked in practice was our 'honest broker' huddle process. We could have spent months perfecting it, only to realize our assumptions were incorrect. Instead, by instituting the huddle early on, we were able to make it a work-in-progress and improve upon it as we gained insight from actual discussions on patient allocation. Another low-tech example involves our methods for evaluation of participant enrollment after Data Mart recommendations are made. We may eventually build a participant attribution workflow module into the Data Mart, but coordinators have been very receptive to using their existing screen logs to supply evaluation metrics whenever requested.

(5) It is constructive to explore and recognize multiple ways to evaluate success. In our early assessment exercises, we recognized that success could be measured by increasing the rate of enrollment for trials and also by reducing the time required by study personnel to evaluate patients for enrollment. Our approach minimizes the need for study teams to forage through medical records and reduces 'cold-calling' by study teams. Given upfront C2C workflow, fewer contacts are needed to enroll the same number of participants and this gain in efficiency has been cited by research teams to be as important as the effective rate of recruitment. We also reduced the need for study teams to go through the provider to gain the patient's permission for research contact. This is especially critical during a pandemic when demands on clinicians' time are already excessive.

(6) Focusing on a specific problem while also considering generalizability for future use cases is advantageous. There is nothing particularly unique about COVID-19 in our approach to building the Data Mart system and we are already thinking about future ways to abstract concepts and technical tools for other recruitment and enrollment workflow solutions.

5. Conclusions and future directions

The REDCap-based Data Mart is operational and yielding beneficial results for the VUMC research enterprise. We will continue to learn and grow the system as we onboard each new study, but the overall framework is working well and equitably serving both patients and researchers in the conduct of research during the current pandemic. While the details of our approach are somewhat unique to VUMC systems, the overall methods and lessons learned should be transferable. This lightweight technology approach is not unique to the COVID-19 recruitment space and we believe it could be customized for use in other biomedical research domains where multiple trials should be considered for each patient (e.g., diabetes clinic, movement disorders clinic). In Appendix 3, we provide details on data sources, logistics, and reporting as guidance for other institutions considering a similar implementation. We are happy to share our model, REDCap project dictionaries, and External Modules with others and plan to do so through webinars and other dissemination channels.

CRediT authorship contribution statement

Tara T. Helmer: Conceptualization, Methodology, Writing - original draft, Project administration. **Adam A. Lewis:** Conceptualization, Methodology, Software, Writing - review & editing. **Mark McEver:** Conceptualization, Methodology, Software. **Francesco Delacqua:** Conceptualization, Methodology, Software, Writing - review & editing. **Cindy L. Pastern:** Methodology, Writing - review & editing, Project administration. **Nan Kennedy:** Writing - original draft. **Terri L. Edwards:** Conceptualization, Writing - review & editing, Supervision. **Beverly O. Woodward:** Conceptualization, Writing - original draft. **Paul A. Harris:** Conceptualization, Methodology, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Rob Taylor for assisting with REDCap integration and validation, Chad Lightner for designing figures for the paper, and Lori Anne Parker-Danley for reviewing and revising the patient messages.

Funding

This work was supported with institutional funds and by the National Institutes of Health [grant numbers UL1TR002243, U24TR001579, contract #75N97019P00279]. The contents of this publication are solely the responsibility of the authors and do not necessarily represent official views of NIH.

Appendix A. . Patient email and text message (English)

English email

Subject: Learning more about COVID-19 research

Hello [first_name],

As you know, COVID-19 has led to a public health crisis. In just a couple of months, the pandemic has made a deep and lasting impact on our patients, our community, our state, our country, and our world.

We need your help—now more than ever. By learning about the COVID-19 research we're doing at Vanderbilt, you can help us win the fight against this pandemic.

As our colleagues in our hospitals and clinics provide the best care possible to all our patients affected by COVID-19, our research community is joining them in the fight. Together with some of the country's top experts in clinical care, infectious diseases, and emergency planning, our world-renowned researchers are focusing on finding ways to:

- best treat all patients who are sick with COVID-19,
- slow or stop the spread of COVID-19 from one person to another, and
- prevent COVID-19 infection.

If you would like to learn more about COVID-19 research at Vanderbilt, tell us how you would like to be contacted. Use the link below to take a short survey.

Click on the link below to open the survey in your web browser:

Patient Preference for Research Contact

If the link above doesn't work, try copying it into your web browser:

[survey-url:patient_preference_for_research_contact]

This link is just for you. Don't forward it to anyone else.

If you have any questions, please call **(615) 322–7343, or email research.support.services@vumc.org**.

Thank you for letting us care for you, and for your support of our research programs.

Sincerely,

Gordon R. Bernard, MD Professor of Medicine Executive Vice President for Research Director, Vanderbilt Institute for Clinical and Translational Research (VICTR)

English Text Message

Hello [first_name],

Vanderbilt Medical Center needs your help in the fight against COVID-19.

If you would like to learn more about COVID-19 research at Vanderbilt, tell us how you would like to be contacted. Please click on the link below to take a short survey.

[survey-url:patient_preference_for_research_contact]

Questions? Call (615) 322–7343 or email research.support.services@vumc.org (mailto:research.support.services@vumc.org).

Thank you for your support of research programs.

Sincerely,

Gordon R. Bernard, MD, Executive Vice President for Research

Appendix B. . Example of study specific logic

Logic for the Adapt Out study

The following logic may be tested using REDCap's built-in reporting feature. DO NOT combine logic from multiple blocks below. Run a report for each individually instead. Mixing logic between multiple repeating instruments sometimes works, but is technically not supported and can cause very odd results.

Include Logic - Match All

```
[donotcontact] != 1
and [is_deceased] != 1
and ([contact] = 1 or [contact_sp] = 1)
and [current_inpatient] = 0
and ([age]>=18)
and
([labs_loinc_code]="94533-7" and [labs_value] = "Detected")
and
(datediff([labs_time], 'now', 'd') <= 7)
```

View Results For Above Logic

Exclude Logic - Match Any

```
([medication_status] = "active")
and
(
  (contains([medication_label], "hydroxychloroquine"))
  or
  (contains([medication_label], "chloroquine"))
  or
  (contains([medication_label], "ivermectin"))
  or
  (contains([medication_label], "remdesivir"))
)
```

View Results For Above Logic

Appendix C. . C2C and Data Mart – Data Sources, logistics Details, and reporting process

- Data Sources
  - The C2C Project
    - A "Test sample collected" message is received from Informatics Aggregation Team (RDW), and a medical record number (MRN) is added to our C2C project
    - Within a few minutes (cron), our COVID-19 Registry module populates patient details and Alerts & Notifications are sent to matching emails or phone numbers
      - Demographics are populated via a FHIR call to Epic
      - Inpatient statuses are populated via a Census Data feed from the Informatics Aggregation Team (RDW)
    - Patients respond at their leisure and update their consent to contact fields, which the COVID-19 Registry module also copies to the Data Mart
  - The Data Mart
    - As new studies are added, they provide their inclusion/exclusion criteria which are manually reviewed & coded into the Advanced Reporting module by our team
    - Every few minutes (cron) our COVID-19 Registry module adds new records to our Data Mart
      - A list of COVID positive MRNs from the Informatics Aggregation Team (RDW) determines which new MRNs to add
      - Inpatient statuses are populated via a census data feed from the Informatics Aggregation Team (RDW)
      - A Data Mart fetch is performed only on new MRNs to immediately populate demographics, labs, medications, allergies, and the problem list
    - Every hour appointments are updated for all existing MRNs via the Appointment Sync module which accesses Health IT's appointment search service (a thin wrapper around Epic's scheduling APIs).
    - Once a day, the Data Mart automatically updates demographics, labs, medications, allergies, and the problem list for all existing records
  - The Study Enrollments Project
    - An E0A rule pushes patient study enrollments in real-time from Epic to the Epic Participant Updater REDCap External Module and Epic Participant Updater (catch all) project
- Resulting Report
  - Every hour, the Advanced Reporting module runs previously defined inclusion/exclusion criteria for all studies against all the above data sources and caches the resulting "Outpatient Study Matches" report on the Data Mart project.

References

[1] K.R. Tuttle, Impact of the COVID-19 pandemic on clinical research, Nature Rev. Nephrol. 16 (10) (2020) 562–564, https://doi.org/10.1038/s41581-020-00336-9.

[2] A. van Dorn, COVID-19 and readjusting clinical trials, Lancet 396 (10250) (2020) 523–524, https://doi.org/10.1016/S0140-6736(20)31787-6.

[3] V. Gewin, Safely conducting essential research in the face of COVID-19, Nature 580 (7804) (2020) 549–550, https://doi.org/10.1038/d41586-020-01027-y.

[4] G. Spiteri, J. Fielding, M. Diercke, et al., First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, Euro Surveill. 25 (9) (2020), https://doi.org/10.2807/1560-7917.ES.2020.25.9.2000178, 24 January to 21 February 2020.

[5] S. Richardson, J.S. Hirsch, M. Narasimhan, et al., Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area, JAMA 323 (20) (2020) 2052, https://doi.org/10.1001/jama.2020.6775.

[6] Burke RM. Symptom Profiles of a Convenience Sample of Patients with COVID-19 — United States, January–April 2020. MMWR Morb Mortal Wkly Rep. 2020;69. https://dx.doi.org/10.15585/mmwr.mm6928a2.

[7] Announcement-New-ICD-code-for-coronavirus-3-18-2020.pdf. Accessed October 14, 2020. https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-3-18-2020.pdf.

[8] D. Doroshow, S. Podolsky, J. Barr, Biomedical Research in Times of Emergency: Lessons From History, Ann. Intern. Med. 173 (4) (2020) 297–299, https://doi.org/10.7326/M20-2076.

[9] Yusuf E, Maiwald M. COVID-19, equipoise and observational studies: a reminder of forgotten issues. Infection. Published online June 25, 2020:1-3. https://dx.doi.org/10.1007/s15010-020-01466-9.

[10] C. Calia, C. Reid, C. Guerra, et al., Ethical challenges in the COVID-19 research context: a toolkit for supporting analysis and resolution, Ethics Behav. (2020) 1–16, https://doi.org/10.1080/10508422.2020.1800469.

[11] P. Soltani, R. Patini, Retracted COVID-19 articles: a side-effect of the hot race to publication, Scientometrics. 125 (1) (2020) 819–822, https://doi.org/10.1007/s11192-020-03661-9.

[12] J.H. Moore, I. Barnett, M.R. Boland, et al., Ideas for how informaticians can get involved with COVID-19 research, BioData Min. 13 (2020), https://doi.org/10.1186/s13040-020-00213-y.

[13] M. Chowkwanyun, A.L. Reed, Racial Health Disparities and Covid-19 — Caution and Context, N. Engl. J. Med. 383 (3) (2020) 201–203, https://doi.org/10.1056/NEJMp2012910.

[14] D.B. Chastain, S.P. Osae, A.F. Henao-Martínez, C. Franco-Paredes, J.S. Chastain, H. N. Young, Racial Disproportionality in Covid Clinical Trials, N. Engl. J. Med. 383 (9) (2020), e59, https://doi.org/10.1056/NEJMp2021971.

[15] Journal JSH and PL| P by BM for TWS. Coronavirus Researchers Compete to Enroll Subjects for Vaccine Tests. Wall Street Journal. https://www.wsj.com/articles/coronavirus-researchers-compete-to-enroll-subjects-for-vaccine-tests-11593968711. Published July 5, 2020. Accessed October 8, 2020.

[16] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J.G. Conde, Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support, J. Biomed. Inform. 42 (2) (2009) 377–381, https://doi.org/10.1016/j.jbi.2008.08.010.

[17] P.A. Harris, R. Taylor, B.L. Minor, et al., The REDCap consortium: Building an international community of software platform partners, J. Biomed. Inform. 95 (2019), 103208, https://doi.org/10.1016/j.jbi.2019.103208.

[18] Overview - FHIR v4.0.1. Accessed October 12, 2020. https://www.hl7.org/fhir/overview.html.

[19] REDCap. Accessed October 14, 2020. https://redcap.vanderbilt.edu/consortium/modules/index.php.

[20] Vanderbilt CIO: When customizing an EHR, don't get "hyper-focused on the tech." Accessed October 14, 20 https://www.beckershospitalreview.com/ehrs/vanderbilt-cio-when-customizing-an-ehr-don-t-get-hyper-focused-on-the-tech.html.

[21] I. Danciu, J.D. Cowan, M. Basford, et al., Secondary use of clinical data: The Vanderbilt approach, J. Biomed. Inform. 52 (2014) 28–35, https://doi.org/10.1016/j.jbi.2014.02.003.

[22] "What is Operation Warp Speed?" | NIH: National Institute of Allergy and Infectious Diseases. Accessed October 14, 2020. http://www.niaid.nih.gov/grants-contracts/what-operation-warp-speed.

[23] Pew Research Center. Demographics of Internet and Home Broadband Usage in the United States. Pew Research Center: Internet, Science & Tech. Accessed October 8, 2020. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/.

[24] Anderson M. Racial and ethnic differences in how people use mobile technology. Pew Research Center. Accessed October 8, 2020. https://www.pewresearch.org/

fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/.

[25] Trial Sites Around the Nation Report Great Interest in Volunteering for COVID-19 Vaccine Trials. Trial Site News. Published August 10, 2020. Accessed October 15, 2020. https://www.trialsitenews.com/trial-sites-around-the-nation-report-great-interest-in-volunteering-for-covid-19-vaccine-trials/.

[26] News PG JoNel Aleccia,Kaiser Health. Thousands Volunteer for COVID-19 Vaccine Study. Scientific American. Accessed October 15, 2020. https://www.scientificamerican.com/article/thousands-volunteer-for-covid-19-vaccine-study/.