UTILIZING BLOCKCHAIN TECHNOLOGY FOR CLINICAL TRIAL OPTIMIZATION

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Yan Zhuang

Dr. Chi-Ren Shyu, Dissertation Supervisor

MAY 2021

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

UTILIZING BLOCKCHAIN TECHNOLOGY FOR CLINICAL TRIAL OPTIMIZATION

Presented by Yan Zhuang,

a candidate for the degree of Doctor of Philosophy, and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Chi-Ren Shyu

_____

Dr. Wei Jiang

_____

Dr. Grant Scott

_____

Dr. Lincoln Sheets

DEDICATION

To

My altruistic parents (Naisheng Zhuang, 庄乃生, and Ming Chu, 初鸣)

For bringing me to the world, giving me a warm home, and raising me right.

My amazing wife (Xiyuan Gao, 高溪远)

For your support, understanding, companion, sharing, patience, encouragement, passion, trust, and endless love.

My lovely daughter (Olivia Zhuang, 庄珺晗)

For letting me know I am not the only one awake during countless research nights.

ACKNOWLEDGEMENTS

First and foremost, I would like to offer my greatest gratitude to my advisor, Dr. Chi-Ren Shyu for all the support and guidance. His dedication, rigorous attitude, enthusiasm for research, and vision guide me to be an independent researcher. I am deeply grateful to have him to be my advisor who has put trust in me and cultivated my potential.

Additionally, I would like to thank my committee members, Dr. Wei Jiang, Dr. Grant Scott, Dr. Lincoln Sheets, for generously offering their time, constructive advice, and help to my study. Special thanks to Dr. Zon-Yin Shae from Asia University for all his support and innovative ideas to help me overcome challenges during my Ph.D. study especially at the early stage of my research. I would also like to thank my current and former colleagues, notably Zainab Al-Taie, Dr. Michael Phinney, Dr. Yuanyuan Shen, Dr. Matt Spencer, from the Interdisciplinary Data Analytics and Search (iDAS) Lab for all their suggestions, help, and encouragements. I also want to thank Mr. Robert Sanders and Ms. Tracy Pickens for their professional assistance over the past five years.

Finally, I would like to thank my friends, foremost among whom are Xiaonan Chen, Justin Floretta, Jingli Gao, Tianhao Shen, Hanqing Sun, Dr. Zhaohui Yang, and Dr. Peng Zhao, for always being there for me through my ups and downs. Final thanks go to my funding sources from the Shumaker Endowment of Biomedical Informatics, Emerging Technologies for Data-Driven Discovery Initiatives, and the Informatics and Data Science Research Initiatives of the University of Missouri.

Table of Contents

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ABI: Application Binary Interface

AI: Artificial Intelligence

API: Application Program Interface

CRF: Case Report Form

CTMS: Clinical Trial Management System

DSMB: The Data and Safety Monitoring Board

EDC: Electronic Data Capture

EHR: Electronic Health Record

eTMF: electronic Trial Master File

FDA: Food and Drug Administration

FHIR: Fast Healthcare Interoperability Resources

GUI: Graphical User Interface

HIE: Health Information Exchange

IoT: Internet of Things

IPFS: InterPlanetary File System

MIMIC: Medical Information Mart for Intensive Care

ONC: Office of the National Coordinator

PHI: Protected Health Information

PoW: Proof of Work

RHIO: Regional Health Information Organization

RLS: Record Locator Service

SDV: Source Data Verification

SEER: Surveillance, Epidemiology, and End Results

TMF: Trial Master File

TPS: Transactions Per Second

VCT: Virtual Clinical Trials

ABSTRACT

Clinical trials are the cornerstone of treatment discovery because they provide comprehensive scientific evidence on the safety, efficacy, and optimal use of therapeutics. However, current clinical trials are facing multiple challenges such as patient recruitment, data capture, and overall management. There are various causes of patient recruitment challenges such as inefficient advertising models, complex protocols, and distant trial sites. Data inconsistency is the main challenge of the data capture process. Source data verification, a standard method used for data monitoring, is resource-intensive that can cost up to 25% of the total budget. The current clinical trial management system market is fragmented and lacks thorough designs with all desired features so that nearly all respondents to management systems from the annual global survey reported dissatisfaction with the current management system. Based on these challenges, disruptive technologies such as blockchain may provide feasible solutions by utilizing its unique features.

Blockchain is an open-source distributed ledger technology that was first applied in the financial sector. Its features such as public audibility, data security, immutability, anonymity, and smart contracts are a good fit for the needs of many healthcare applications. However, there are several common challenges of blockchain technology so that most blockchain designs for healthcare applications are still in the early stage of implementation.

This dissertation aims at optimizing clinical trials by developing multiple applications using blockchain technology to provide feasible solutions to the current challenges. We will use real-world data to conduct large-scale simulations to evaluate the feasibility and performance of proposed blockchain models for clinical trial applications.

CHAPTER ONE - INTRODUCTION

1.1 Background

Clinical trials are considered the cornerstone of the development of new drugs or treatments because they have investigated the safety and efficacy of the new therapeutics using a standard protocol [1]. However, current clinical trials are facing multiple challenges related to patient recruitment, persistent monitoring, and management. 86% of clinical trials don't achieve their recruitment goals on time and 19% of registered clinical trials have been either closed or terminated due to failure to reach recruitment goals [2]. Failure to meet recruitment goals on time results in incomprehensible statistical results, premature trial termination, and delay of the study timeline that could double the planned recruitment period [3]. It can cost up to 8 million dollars for each day of delay [4]. Possible causes for patient recruitment issues are related to inefficient advertising models [16], concerns about the legitimacy of the clinical trials [3], unreachable clinical trial sites that deter potential subjects with frequent travels to the site during the trial [16], etc. The challenges of monitoring mainly surround data collection such as data security concerns [5, 6], data inconsistency caused by human error or falsification [7], and difficulties of data exchange across multiple healthcare facilities [8]. Source data verification (SDV), a verification of the conformity of the reported data with source data, is costly with an average of 25% of the entire clinical trial budget [9]. The Clinical Trial Management System (CTMS) is a set of software tools used for managing clinical trial processes. A complete design of CTMS needs to be secure, cost-efficient, regulation compliant, traceable, and auditable to manage the process for each phase of the study [10-12]. However, the current CTMS market is fragmented and lacks thorough designs with all needed features and management tools [12,

13]. According to the 2019 Unified Clinical Operations Survey provided by Veeva, a global life-science service, nearly all respondents (99%) had issues with their current CTMS and 90% of the respondents reported a significant deficiency in at least one CTMS application [14]. With these challenges, emerging technologies, such as blockchain, may provide potential solutions for needed functions in clinical trials to tackle the aforementioned challenges.

Blockchain is a distributed ledger technology which has been first applied in the financial sector [15, 16]. The success of the Bitcoin cryptocurrency, which is one of the blockchain's most popular applications, shows the robustness, security, and consensus mechanism of the blockchain system [17]. Blockchain also has other features such as anonymity, decentralization, immutability, data provenance, and public auditability [16]. All the transactions that have occurred in the blockchain are distributively stored into each node which can be any active electronic device without specific hardware requirements inside the blockchain system [18]. Any transaction needs to be validated by the users in the blockchain before it is written into the system [19]. Since the system is fully decentralized, all the transactions can be audited publicly by all the users. The public auditability feature can solve many clinical trial challenges such as providing an oversight role of FDA, promising legitimacy of the clinical trials, and making users aware of the available trials. Once the record is stored in the blockchain, it cannot be modified [20]. This immutability feature ensures data consistency and provides trust to users. With the smart contract feature which is a self-executing, coded protocol agreed between senders and receivers [21], blockchain can solve more complex challenges that will be elaborated in each aim.

Despite the features of blockchain fitting most healthcare applications, there are several common challenges of the most current blockchain models for healthcare applications: (1) inadequate public/private key management systems; inability to retrieve any information once a user has lost the private key [22], (2) loss of privacy caused by the transparency of the distributed system [23, 24]; the ability of all the users in the blockchain to view all the data stored in the blockchain, (3) scalability constraints, considering Ethereum can handle roughly 15 transactions per second [24, 25], and (4) most blockchain applications in the healthcare area are still in the design stage and have not yet been implemented [26]. A practical blockchain framework for healthcare applications is needed to evaluate the feasibility of utilizing blockchain technology for the healthcare field.

1.2 Study objectives

This study aims at utilizing blockchain technology to provide multiple practical problem-oriented blockchain models for clinical trial optimization. We have achieved this objective by pursuing the following specific aims.

Aim 1. We have built a generalized blockchain architecture that provides data coordination functions, including data requests, permission granting, data exchange, and usage tracking, for various healthcare application developments. Developers do not need extensive experience in blockchain to implement applications on the proposed blockchain architecture.

Aim 2. We have built a blockchain architecture to achieve a patient-centric Health Information Exchange (HIE) which provides patients full control of their health records by personalizing data segmentation to regulate sharable information and authorizing “allowed list” for clinicians to access their data. Timely HIE can improve the healthcare outcomes

as well as the quality and efficiency of the clinical research process such as patient recruitment. Large-scale simulations have been conducted to test the feasibility of sharing confidential information using blockchain technology and the performance of the proposed architecture.

Aim 3. We have investigated the current clinical process and implemented several blockchain models to tackle issues related to patient recruitment, patient engagement, data capture, persistent monitoring, and overall clinical trial management. We have also built a blockchain architecture to achieve an innovative approach to conducting clinical trials solely through a digital platform.

1.3 Dissertation organization

In this dissertation, we have provided multiple blockchain models fit for the needs of different clinical trial applications. Each proposed model has been tested for feasibility, stability, and scalability by conducting simulations. In each chapter, we have elaborated on the rationale of using blockchain technology to solve the current challenges. Chapter two gives an introduction of blockchain technology in detail along with a description of all its unique features. The subsequent chapters have applied many blockchain concepts without repeating the explanation. Chapter three describes a generalized blockchain architecture that provides data coordination functions for a wide spectrum of healthcare applications not limited to clinical trial application developments. The architecture offers health technology community blockchain features for application development without requiring developers to have extensive experience with blockchain technology. In chapter four, a blockchain model is built to achieve a patient-centric Health Information Exchange (HIE). A large-scale simulation of this patient-centric HIE process was performed, and the model's

feasibility, stability, security, and robustness were quantitatively evaluated. Chapter five presents a blockchain model to enhance clinical trial recruitment. The model contains multiple trial-based contracts for trial management and patient engagement and a master smart contract for automated subject matching, patient recruitment, and trial-based contracts management. Chapter six illustrates a comprehensive clinical trial management system implemented using blockchain. Multiple applications are designed through smart contracts that span all stages of clinical trials. Chapter seven describes a blockchain framework to achieve virtual clinical trials, an innovative approach to conduct home-based rather than site-based clinical trials. This chapter brings the healthcare community a fresh idea of conducting clinical trials by utilizing blockchain's unique features. Chapter eight concludes by reviewing contributions presented in the prior chapters, and by discussing the limitations and future work.

CHAPTER TWO - BLOCKCHAIN TECHNOLOGY

This chapter adopts partial contents with minor modifications from our previous publication in AMIA Annual Symposium proceedings in 2018 [27].

Zhuang Y, Sheets L, Shae Z, Tsai JJP, Shyu CR. Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials. AMIA Annu Symp Proc. 2018 Dec 5;2018:1167-1175. PMID: 30815159; PMCID: PMC6371378.

2.1 Introduction

Blockchain is a distributed ledger technology that was first applied in the financial sector [16]. Bitcoin is one of the most popular applications of blockchain that shows its security, durability, and robustness. Blockchain is an open-source platform to allow all users to make transactions without a mediating party. It reduces the cost of transactions and the time of working with third parties. The entirety of the transaction and validation processes are performed by users inside the blockchain. When a user makes a transaction, all of the information from this transaction is encrypted using cryptographic algorithms and broadcast to every user in the blockchain network for validation [16].

Validation processes contain two parts. The first part is the validation of the user's key pair; the second part is validation that the user's account balance is sufficient to make the transaction [16]. Each user has a unique key pair consisting of a public key and a private key. The public key is similar to a proxy user ID in the blockchain system so that no blockchain node, which can be any electronic device that can install the blockchain system, is able to know the patient's identification. The private key is similar to the user's signature. Each transaction is sent to the receiver's blockchain account derived from the public key

and digitally signed by the sender's private key. The receiver needs to validate the identity of the sender by checking whether or not the public key matches the sender's digital signature. A user can create a username and password or use biometric information that is mapped to the public and private key instead of memorizing the real key's value. Since we use Smart Contract to regulate all transactions executed in the blockchain, transactions only need to be validated through the confirmation of the sender's identity. If there is a consensus of most users in the system, this transaction is written into the coming block. All validated transactions occurring after a previously created block are recorded in the next block. All transactions are secure, trusted, auditable, and immutable.

Blockchain can be set up as a "public chain" or a "private chain." A public chain is also known as a "permissionless" chain, which means anyone can join this chain and see all the transactions which have occurred since its beginning [28]. They can also participate in the validation and consensus process. Bitcoin is the cryptocurrency inside this public blockchain. A public chain is a fully distributed chain, which means that all transactions are dependent on a consensus decision of all nodes. The stability of blockchain depends on a mass of participating nodes; they contribute computing power to ensure the reliability of the consensus. A private chain is also known as a "permission chain" [28]. Users must get permission to join the private chain. Each node installs the specific "genesis block" of a private chain to join the system. The private chain is not fully decentralized since the creator of a private chain decides who has permission to join this chain [29]. The great benefit of private chains versus public chains is it is easy to regulate the users and transactions to ensure privacy, scalability, and security. For EHR data, making transactions visible only to authorized users is ideal. In this work, we utilize a private blockchain system

for more security and ease of regulation. Each clinical trial could have different regulations based on their protocols, and each hospital could have different regulations based on their policies. Therefore, a private chain is better than a public chain to offer customized functionality for HIE and clinical trial settings.

Every blockchain will start with a genesis block which is the first block in the chain. We can set up multiple parameters in the genesis block which determines the characteristics of the private chain. A parameter called the "gas limit" restricts the transaction size. In the Ethereum blockchain system, each bytecode inside a transaction has a pre-defined "gas amount". When users deploy Smart Contract in a system, they must pay a "gas fee" for the deployment. If the gas fee exceeds the gas limit, the transaction will be declined. In our implementation, which does not focus on the financial aspects of blockchains, we have assigned a sufficient balance for each user inside the private chain so that every user will be able to send transactions at any time. We have also set up a maximum value of the "gas limit" in case of some clinical information is too large to be sent through the blockchain. Another important parameter is called difficulty, which determines the "block generating rate" [29]. Private chains can set up a high generating speed that can support real-time transactions. In addition, private chains also have Smart Contract protocols to regulate transactions and validate users' identities through coding the policies of different clinical sites and the protocols under different scenarios.

There is a debate regarding private chain versus databases since a private chain is not a fully distributed system [30]. Shared databases could share "read" permissions with multiple authorized users without any issues. However, when users get shared "write" permissions, they can easily modify the master file which could result in unrecoverable

errors [31]. Private blockchains provide more secure identity validation and a higher level of error checking than regularly shared databases. Private chains can code any permission level for any user. When there is a conflict between users' requests and protocols, Smart Contract uses cryptographic algorithms to ensure that invalid transactions are not added to the blockchain. Even if a hostile attack on a user's account sends malicious requests, each user keeps a copy of the transaction history and can recover the system from any given timestamp [19]. Private chains can create peer-to-peer networks; any authorized party inside the private blockchain can query transactions from another party without changing the original data.

2.2 Smart Contract

Smart Contract is an agreed computing protocol on top of the blockchain, used especially in the Ethereum blockchain system. It was first proposed by Nick Szabo in 1994 to allow distributed ledger systems to regulate contracts [32]. These contracts could be coded as computing protocols, stored inside blockchain systems, and self-executed. Ethereum is a distributed ledger that runs Smart Contract [29]. Smart Contract is written using Solidity, which is a Turing-complete language and is expected to encode all rules needed for HIE. It can also contain data exchange policies from clinical sites.

Smart Contract is compiled using a Solidity compiler residing on a blockchain. After deploying Smart Contracts to the blockchain, the system returns an address [33]. Users who want to use other Smart Contract functions must use the application binary interface (ABI) of Smart Contract and Smart Contract address. Users can only check the ABIs for a different Smart Contract. The source code is anonymous and is believed unhackable [34].

```
pragma solidity ^0.4.13;

contract Ownable {
  address public owner = msg.sender;
  /// @notice check if the caller is the owner of the contract

  modifier onlyOwner {
    require (msg.sender == owner) ;
    _;
  }
  address[] pharmas;
  function Add_pharmas(address[] pharmas_) public
  onlyOwner
  {
      for (uint i = 0; i < pharmas_.length; i++) {
        pharmas.push(pharmas_[i]);
      }
  }

  mapping (address => uint) perms;
  function set_permission() public{
      for (uint i=0;i<subjects.length;i++)
      {
          perms[subjects[i]]=3;
      }
      for (i=0;i<pharmas.length;i++)
      {
          perms[pharmas[i]]=2;
      }
      perms[owner]=1;
      //1 is highest, 2 is high, 3 is low
  }
```

Figure 1. A smart contract example to demonstrate ownership and permission levels of different roles inside the blockchain system

Smart Contract is a coded consensus protocol: all the users in a blockchain system must follow the protocols to make transactions. Figure 1 is an example of a Smart Contract showing how different permission levels are set up for different users.

The contract is deployed by the owner of the contract. In our system, we hypothesize and assign the FDA as the owner of the Smart Contract. As shown in Figure 1, the "onlyOwner" modifier (line 7) ensures that functions with this modifier can only be executed by the owner. The "Add_pharmas" function (lines 12-18) with the "onlyOwner" modifier ensures that only the owner can add clinical sponsors; executions of this function from other users would be declined automatically. Different roles in the system have different privileges to execute different functions. The "set_permission" function (lines 21-32) is a public function: any user in the system can call this function to initialize their own permissions. There is a pre-defined permission level giving the FDA the highest privilege.

2.3 Blockchain adapter

In our system design, we have converted blockchain nodes into blockchain adapters that abide by the local health IT policies [35]. These nodes need to take the following steps to build a "blockchain adapter" to communicate with the system: (1) deploy the correct "Genesis block" (the starting block of the blockchain), which is a JSON file containing the blockchain's unique characteristics, and add the starting node as a peer node; (2) build a remote procedure call server which can communicate with servers outside the adapter and secure EHR databases inside the healthcare facility's firewall; and (3) build a temporary receiving database outside the firewall to store data received from all other healthcare facilities' blockchain adapters. These steps are embedded into an installation file that can automatically convert a node into a blockchain adapter. Once a blockchain adapter joins the blockchain system, a pair of keys are automatically generated for the adapter and managed by the local administrator.

2.4 Assumptions

To conduct simulations using the proposed blockchain models, we made the following assumptions: (1) Each participating site, including the sponsor, trial sites, site institutional review boards, and the Food and Drug Administration (FDA) is required to provide at least one blockchain adapter, (2) each participating site's local health IT has agreed on the adapter setting, (3) patients have authorized the blockchain system as well as the application to access their health records, (4) all participating parties have an administrator to operate the system, and (5) all users can operate the system properly.

CHAPTER THREE - GENERALIZABLE BLOCKCHAIN ARCHITECTURE

We have published this chapter in the Journal of Medical Informatics Research in 2020 [36]. This chapter adopts its main contents with minor modifications.

Zhuang Y, Chen YW, Shae ZY, Shyu CR. Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation. J Med Internet Res. 2020 Jul 27;22(7):e19029. doi: 10.2196/19029. PMID: 32716300; PMCID: PMC7418010.

3.1 Background

The healthcare industry generates abundant health data from various sources [37]. The increasing adoption of digitalized healthcare records such as Electronic Health Records (EHRs) provides the opportunity for healthcare data analytics and the coordination of results with patients' care [38, 39]. Many clinical research applications need to be designed by maximizing the potential benefits of EHR usage such as supporting drug development [40], expediting the recruitment process [41], and improve the clinical trial outcomes, [42, 43]. However, multiple barriers to data coordination exist: (1) data privacy and security concerns during HIE [44, 45]; (2) the limitations of institutional privacy rules [38, 46]; and (3) the time-consuming process of generating agreements on data exchange between institutions [47, 48]. There are security and privacy concerns about the exchange of sensitive health data [45]. Due to the Health Insurance Portability and Accountability Act (HIPAA), legislation limits EHR access without patient authorization [38, 49]. Therefore, there needs to be a sustainable and secure data collaboration mechanism by

which each data owner can maintain control of their data, and only if the owner of the data allows that to occur [50].

Facing the challenges of data coordination, blockchain is considered to be a disruptive technology that can potentially provide a solution [24, 26]. There have been many efforts to apply blockchain to areas of healthcare [15, 51, 52]. However, most blockchain applications in the healthcare area are still in the early stages of implementation [26]. In this chapter, we will describe a generalized layered architecture that fits a wide spectrum of healthcare applications not limits to clinical research with an essential characteristic as cross-site data coordination [36]. The layered architecture provides a blockchain platform with predefined functions for data collection for developers to implement healthcare applications including clinical research applications without extensive experience in blockchain.

3.2 Methods

The Blockchain network offers advantages in managing digital assets [53, 54]. The well-known digital asset managed by blockchain is Bitcoin or, in general, cryptocurrency [15]. In healthcare applications, EHR access is the digital asset in management. The health data for patients who opt to participate in the blockchain are encrypted and stored in secured off-blockchain databases located in healthcare facilities protected by their own firewalls. The foundation private blockchain is used to store all transactions referring to EHR requests and exchanges and the metadata that contain pertinent healthcare data with the identifications of patients and healthcare facilities. In addition, the following components are captured in the metadata: time of creation, dataset location, access permission and control, data decryption, and data authenticity.

As blockchain is a fully distributed system, we have built three layers on the top of the private blockchain network: transaction layer, interfacing layer, and application layer. As shown in Figure 2, the transaction layer consists of two types of smart contracts coded in Solidity, namely, *EHR manager* smart contracts, and *user manager* smart contracts, to manage the storage of and access to metadata that are encrypted with decentralized validity and authenticity checks using blockchain security [55]. These two smart contracts are fixed in the system and are not permitted to change.

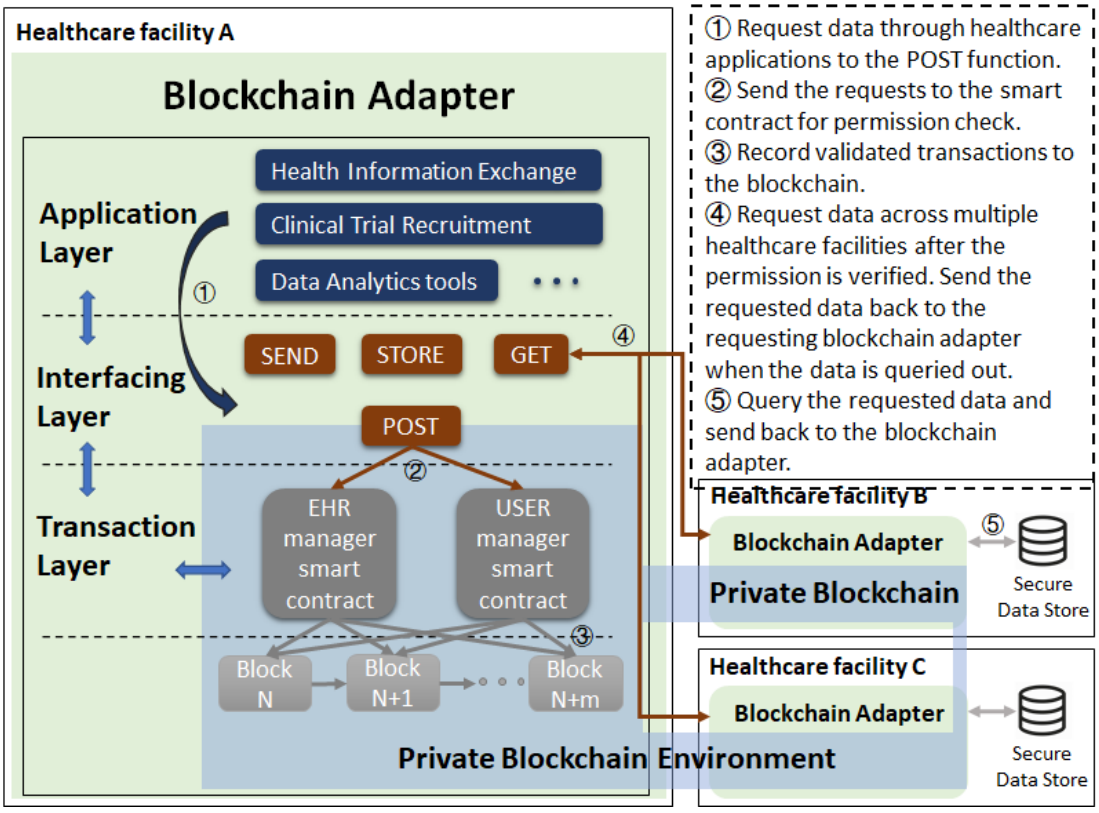

Figure 2. Overall layered blockchain architecture. The transaction layer consists of two smart contracts to manage data access tasks. The interfacing layer interacts with the blockchain environment, graphical user interfaces, and other blockchain adapters. The application layer provides a flexible platform for healthcare application development. The figure shows the general process of data requests using the architecture.

Data requests from one healthcare facility to another trigger certain functions of smart contracts through the interfacing layer. Only through the information in the trustable metadata can original data be retrieved and verified for authenticity by the interfacing layer. Applications such as subject recruitment for clinical trials, EHR management, and artificial intelligence (AI)–based data analytics tools can be built on the application layer. This layered architecture has the following benefits compared with the previous blockchain systems used for healthcare applications: (1) compatibility of most healthcare applications that require data exchange, (2) semi-publicity to fix the blockchain settings and smart contract functions but retain most blockchain features, (3) security settings of each layer to protect the identities and data of patients during an exchange, and (4) traceability of who have accessed the data and how they used the data.

3.2.1 Environment Setup

To build the blockchain system, each healthcare facility is required to provide at least one blockchain adapter [56]. To ensure the security of healthcare data and meet the needs of current EHR operations, our blockchain system does not store patient data. There are two main reasons for this. First, it is not practical to store a large chunk of healthcare data in a blockchain because of the healthcare facilities' policies of sharing health information and blockchain storage constraints [15]. Second, the healthcare industry is still unreceptive to allowing patient data to move across the blockchain network [49]. A metadata set containing pertinent information of the original EHR data is created and submitted to the blockchain platform. The creation and updating of metadata are recorded into a chain of data blocks in the blockchain. These transactions executed via smart contracts are immutable and traceable, thus creating a trustable metadata transaction. As

there are different interoperability standards, such as Fast Healthcare Interoperability Resources (FHIR) and Health Level 7 version 3, there will be different metadata points of different data stores for the receiver to choose from. The receiver can choose the compatible interoperability standard of their home department's standard during the data coordination process [57, 58]. The metadata owner, who is the same as the dataset owner, can grant, reject, or revoke access permission automatically via smart contract or interactively by means of electronic notification and confirmation. For example, smart contracts can be programmed to grant or reject access permission based on time or data type or to delegate access permission to a specific user. In all cases, because a third-party intervention is not necessary for granting or rejecting permissions, the time efficiency of data sharing can be greatly improved.

The patients need to go to the healthcare facilities to opt-in to the system so that they can claim ownership of their data. The metadata permission control carried out by smart contract is anonymous, which ensures privacy. The metadata used for locating encrypted data are communicated with the secured data stores via the HTTPS protocol, and the result is communicated back to the user via the same protocol; thus, it is considered to be a secure data transfer.

3.3.2 Foundation Private Blockchain network

The foundation of our layered architecture is a private Ethereum blockchain that involves an immutable chain of data blocks consisting of committed ledgers, and multiple blockchain nodes synchronously maintaining the same chain of data blocks. In the overall architecture, this layer ensures data immutability, decentralized consensus, data transparency, and traceability. The private blockchain is initiated from a starting node with

special settings to make the blockchain unique. The smart contracts will be deployed through the starting node when the private blockchain is built. All the participating nodes from healthcare facilities must obtain permission from the starting node to join the system. This procedure will disallow unauthorized parties from joining the system. As the participating nodes are joined into the system, the blockchain will automatically generate accounts for their blockchain adapters. All the other users such as patients, healthcare providers, etc. need to opt-in to the system through healthcare facilities. The blockchain accounts will be generated for the users from each healthcare facility as soon as the applicants' identities are proved.

The private blockchain stores all the transactions for (1) patients and healthcare facilities granting, revoking, and denying access to their EHR data; (2) authentication of patients and healthcare providers to retrieve the EHR data; and (3) healthcare facilities to store metadata for patients' visits. The transactions will record the receiver, sender, contained data, and the timestamp into the blocks through blockchain adapters. Users can also make transactions in the backend blockchain console through the blockchain node. These transactions still need to pass smart contract regulations to become effective. Most users will interact with the Graphical User Interfaces (GUI) built on the *Application Layer* to execute functions in the blockchain system.

3.3.3 Transaction Layer

The transaction layer consists of two smart contracts that specify a metadata model for healthcare records and methods that regulate data access rights, permission policies, and data encryption. Two smart contracts, the *EHR manager* smart contract and the *user manager* smart contract, are deployed to the blockchain to securely accomplish the basic

EHR management tasks. The *EHR manager* smart contract can only be used by healthcare facilities to submit the EHR metadata to the blockchain. The user manager smart contract is used by patients or facilities to manage access to their data. Once a patient has opted into the system from a healthcare facility, the healthcare facility's blockchain adapter will automatically encrypt his or her patient ID and public key using the blockchain adapter's public key and input to the user manager smart contract. Healthcare facilities will have adapter IDs stored in the user manager smart contract.

3.3.3.1 EHR manager smart contract for submitting metadata to the blockchain

The *EHR manager* smart contract (as shown in Figure 3) defines several structures to record patient information: EHRDataID and EHRdata define the metadata components, PatientID stores the patient ID and healthcare facility ID for the registered patients, and patientData maps the different healthcare facility visit records of patients with the patients' IDs. Once the blockchain adapter receives a record from the EHR system, the blockchain adapter automatically performs the following steps to submit the metadata to the blockchain through an *addEHR* function: (1) extract the patient ID from the EHR dataset, (2) find a public key associated with the patient from the *user manager* smart contract, (3) generate a random *data key* for encrypting the EHR dataset, (4) encrypt the EHR dataset using data key and store the encrypted data to an *off-chain* secured data store, (5) use the patient's public key and adapter's private key to encrypt the data key. Call the encrypted data key "key cipher.", and (6) submit the following metadata to the blockchain: patient ID, encrypted dataset location as a URL, key cipher, and associated blockchain adapter ID.

The blockchain adapters will generate public and private key pairs following the Diffie-Hellman protocol between the local hospital adapter and the patient using Node.js.

The key cipher makes the data key more secure and can only be computed when the remote hospital adapter's account is validated and with the patient's authentication. Figure 4 shows one patient's decoded metadata retrieved through blockchain using the Remix integrated development environment, which is an open-source visualization tool used for interacting with blockchain nodes and smart contract development and deployment.

```
pragma solidity ^0.4.13;
contract EHRMgrContract {

struct EHRDataID {
    bytes32 datasetID;      // Dataset ID
    string datasetDesp;     // Dataset description
    bytes32 timeOfCreation;     // dataset creation time
}

struct patientID {
    bytes32 patient_ID;     // patient ID in a hospital
    bytes32 hospital;       // hospital ID
    bytes32 patientName;        // patient name
}

struct EHRData {
    EHRDataID   ehrdataID;  // dataset ID struct
    string datasetLoc;      // an URL or a data object hash
    bytes32 keycipher;      // cipher of the encryption key for EHR
    address bcAdapter;      // blockchain adapter address
}
struct patientData {
    patientID patientid;        // patientID in a hospital
    mapping (bytes32 => EHRData) ehrCollection;    // all EHRs for a patient in a hospital
}

mapping (address => patientData ) patientDataSet; // user info for the EHR datasets

function addPatient(bytes32 _patient_ID, bytes32 _hospital, bytes32 _patientName, address _patient) public {
    patientDataSet[_patient].patientid.patient_ID=_patient_ID;
    patientDataSet[_patient].patientid.hospital=_hospital;
    patientDataSet[_patient].patientid.patientName=_patientName;
}// to add a patient to blockchain

function addEHR(string _loc, bytes32 _keycipher, bytes32 _datasetID, string _datasetDesp, bytes32 _timeOfCreation, address _patient) public{
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.datasetID=_datasetID;
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.datasetDesp=_datasetDesp;
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.timeOfCreation=_timeOfCreation;
     patientDataSet[_patient].ehrCollection[_datasetID].datasetLoc=_loc;
     patientDataSet[_patient].ehrCollection[_datasetID].keycipher=_keycipher;
     patientDataSet[_patient].ehrCollection[_datasetID].bcAdapter=msg.sender;
}// to submit an EHR metadata to blockchain

}
```

Figure 3. The main part of the EHR manager smart contract code defines the Metadata structure. Blockchain adapters must extract the information and calculate the encrypted keys then store them into the smart contract. The record will automatically associate with the adapter's Blockchain ID.

transaction hash	call0x9a095c6a3c41390172232fa6e2441bb07120fd7a0x21966efeaa8e307a aa22d819639abf8b8fd1a4520x0a79309b0000000000000000000000009a095c 6a3c41390172232fa6e2441bb07120fd7a
from	0x9a095c6a3c41390172232fa6e2441bb07120fd7a
to	EHRMgrContract.retrieve(address) 0x21966efeaa8e307aaa22d819639ab f8b8fd1a452
hash	call0x9a095c6a3c41390172232fa6e2441bb07120fd7a0x21966efeaa8e307a aa22d819639abf8b8fd1a4520x0a79309b0000000000000000000000009a095c 6a3c41390172232fa6e2441bb07120fd7a
input	0x0a7...0fd7a
decoded input	{ "address _patient": "0x9A095C6a3c41390172232Fa6e2441BB07 120fD7A" }
decoded output	{ "0": "string: mongo --host 128.206.20.167 -d tmp --port 27017 --username 39791 "1": "string: e0697b0c5c0dcebc95", "2": "string: 08838", "3": "string: Diagnosis:C508", "4": "string: 2019-07-28" }

Figure 4. Example of a patient's metadata retrieved through blockchain. The information is decoded by the Remix web-based integrated development environment, which is connected to the retrieving blockchain node. The patient's metadata contains data location, key cipher, dataset ID, dataset description, and creation time.

3.3.3.2 User Manager Smart Contract for the Health Information Exchange Process

A scenario is described in this section to show the process of healthcare provider X retrieving patient A's EHR. Through a mobile application or web browser with biometric authentication, Patient A can grant access privileges to healthcare provider X using the application program interface (API) to set data permission on the *interface layer*. In some cases, more than one clinician is involved in the patient's care. The healthcare facility needs to create a shared blockchain account for the provider's department so that all involved clinicians can access the patient's data with one-time authentication [59]. Blockchain adapters will record who has accessed the data and submit it to the blockchain. The

blockchain adapters from the receiver's home healthcare facility will perform the following steps for the data retrieval process:

1. Verify healthcare provider X's permission to access patient A's records through the *EHR manager* smart contracts.
2. Retrieve patient A's metadata from the *EHR manager* smart contract.
3. Request the encrypted EHR dataset from the remote healthcare facility via an HTTPS service provided by a blockchain adapter.
4. Retrieve the encrypted data using encrypted dataset location information in EHR metadata.
5. Decrypt the EHR dataset. This step involves decrypting the key cipher by patients to obtain the data key that decrypts the EHR dataset.

Similar processes will be used for sharing data between healthcare facilities and for patients retrieving their own EHR records. The entire process will be performed automatically through the blockchain adapter.

3.3.4 Interfacing Layer

The *interfacing layer* provides 4 high-level methods: *get* the healthcare data from different facilities, *store* the encrypted data securely, *post* metadata or data request to the blockchain via smart contracts in the *transaction layer*, and *send* the encrypted data to the receiver who has gained permission from the data owner. Using the functions in this layer, application developers can implement distributed data applications (DApps) without the knowledge of smart contracts and the underlying blockchain network. This layer consists of APIs and HTTPS web services to define a set of primitive coordinate functions: (1) submit data, (2) set data permission, and (3) retrieve data. The data submission API will

extract metadata from the original data and call the *transaction layer*'s smart contracts to record it to the blockchain. It will also encrypt the original data and store the encrypted version in a secure off-blockchain data store. The data retrieval API will call smart contracts to retrieve metadata from the blockchain, verify encrypted data authenticity with metadata, and decrypt the encrypted data in the off-blockchain data store to obtain the original data. The data permission setting API will call the *transaction layer*'s smart contracts to set access policies and methods for a piece of metadata. Information contained in metadata is used to retrieve and decrypt data. The HTTPS web services provide secure data transport when data are to be transported through a channel and can potentially be eavesdropped on. Using the blockchain adapter to serve as a gateway to the EHR system minimizes the concerns of data exchange security.

We implemented blockchain adapters as a Node.js application and used the web3.js package for interfacing with a blockchain node and https.js package for HTTPS secure web services. web3.js is also available in the Python library as web3.py. The HTTP-based web services are mainly used for communication among blockchain adapters. The blockchain adapter is embedded as software that will install the missing component automatically, such as node.js and web3.js. Figure 5 shows a high-level block diagram of a blockchain adapter. The metadata extractor extracts metadata such as patient ID and dataset ID from the EHR dataset for data identification purposes in the blockchain. The data and patient ID manager maps the patient ID to the dataset ID and records the information in both the *user manager* smart contract as well as the *EHR manager* smart contract. The dataset encryption block in Figure 5 encrypts the EHR dataset and stores it away in the secure data store with a URL or a data object hash for future access.

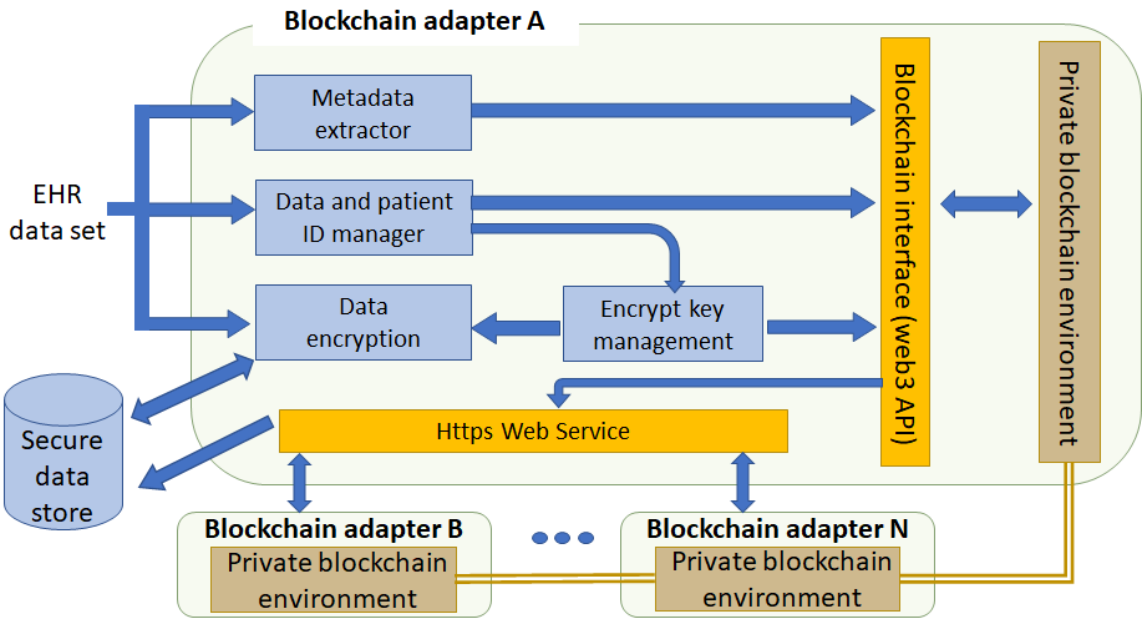

Figure 5. Blockchain adapter components and functions. Blockchain adapter extracts the metadata from the electronic health record, encrypts the electronic health record, stores the encrypted electronic health record into the secured data store, and maps the patient ID and data ID to the blockchain account. Blockchain adapters use HTTPS to interact with other adapters and communicate with the foundation private blockchain network through a blockchain interface.

From the security and reliability point of view, the following design guidelines are strictly followed for the blockchain adapter:

1. Blockchain adapter is modeled as a nonhuman blockchain user and has its own private and public key pair when the blockchain account or address is established. The public key is made public via the *user manager* smart contract. The private key is kept in a blockchain adapter.

2. The HTTPS service uses a separate key and certificate file.

3. The Diffie-Hellman key agreement protocol is used for dataset encryption.

4. When one blockchain adapter fails, transactions (dataset exchange) with the associated organization will be interrupted, but transactions among all other organizations will not be affected.

3.3.5 Application Layer

With the above *Interfacing Layer* architecture and smart contract setup, many healthcare applications involving data exchange can be developed in the *application layer*. This layer will rely on the *interfacing layer* to securely collect the data and then perform data analytics. Applications will not change the existing blockchain settings. These applications can allow researchers or data owners to have better use of the EHR data. For example, personal health records management can be developed on the *application layer*. A patient's identity will be verified through smart contracts in the blockchain. All patient records can be retrieved through the *interfacing layer*. In addition to the HIE application, subject recruitment for clinical trials could also be developed in this layer. Clinical trial sponsors need to obtain permission from the patients through blockchain adapters from clinical trial sites before the matching process [60]. After the patients grant the sponsors permission to gather their data, clinical trial sponsors can use the data analytics tools developed in the *application layer* to match the patients with their recruitment criteria automatically. We implemented these two sample applications on our private blockchain system. The Results section shows the interactions of the *application layer* and the blockchain system and the simulation results.

3.4 Case study

To test the feasibility of our layered architecture, we built a blockchain environment that contains one starting node and four healthcare facility nodes. A blockchain adapter has been installed to each node to communicate with the blockchain and its own secured data store, which was implemented using MongoDB. We created 100 accounts for healthcare providers and 10,000 patient accounts on each healthcare facility node of patients' records

from the Surveillance, Epidemiology, and End Results (SEER) dataset. A total of 2431, 2587, and 2505 patients have multiple records distributivity stored in 2, 3, and 4 healthcare facility nodes, respectively. The remaining patients' records were stored in a single facility node. Selected records were stored in each secured data store using an automated script following the same procedure described in the *EHR manager* smart contract. The ownerships have been claimed when the metadata was pushed into the blockchain using the smart contract in the *transaction layer*. After setting up the environment, we implemented an application as HIE to interact with our blockchain system. The application is an example of the potential use of layered blockchain architecture. We tested the accuracy, scalability, and speed of our system

This application provides an interface for users to manage access to personal healthcare records through the *Transaction Layer*. Patients can use this application to grant and revoke access to their records. Patients can also track how many times their records have been accessed through this application. To test the accuracy, speed, and scalability of our system, we simulated the process of patients granting permission to healthcare providers of their EHR. We developed five scripts to automate the simulation process by (1) randomly selecting one patient to grant one healthcare provider permission to their EHR per second for an hour and recording the timestamp, (2) recording the timestamp when the healthcare providers received the permission, (3) recording the timestamp when the healthcare providers received the data, and (5) adding one patient per second to script one then repeating scripts #1 to #4 until reaching the system limitation because of the known scalability constraints of Ethereum [25].

The simulation only contains the period of interactions with the blockchain. Retrieving data is an off-chain process through the HTTPS portal and varies with different healthcare facilities. From our simulation results, the system breaks at a certain point when the scale reaches 14 transactions per second (TPS). We simulated 331,142 access-granting transactions. All the transactions have successfully retrieved the records except the last second's 14 transactions due to reaching the Ethereum scalability limitation. The average time of writing a transaction to a block is 11.271 (SD 2.208) seconds. We did not find a correlation between TPS and validation time. All healthcare providers received the metadata in an average of 1.73 seconds.

In this study, the scalability of the blockchain using various transaction frequencies from 1 to 14 TPS through blockchain adapters was tested. Figure 6 shows the time spent granting permission from different scales (the 14 TPS group was excluded because of incomplete results). Once the permission is granted by writing the transactions into blocks, the receiver can retrieve the metadata from the smart contract through the blockchain adapter without making another transaction for users to validate the legitimacy. This means that the average time of receiving metadata is much shorter than the grant permission. The script of the 9 TPS group runs slowly compared with the former groups. All the blockchain nodes were restarted separately, and the script was restarted with the 10 TPS group. The speed is affected by the processing speed of the blockchain nodes and Ethereum performance. The starting node's blockchain adapter was used to control the overall frequency. All transactions from the blockchain adapters of healthcare facility nodes will queue in the starting node's adapter until the earlier batches of transactions have been

executed by each blockchain adapter. We controlled the overall frequency as 13 TPS, which avoids Ethereum's scalability constraints by spacing the transactions.

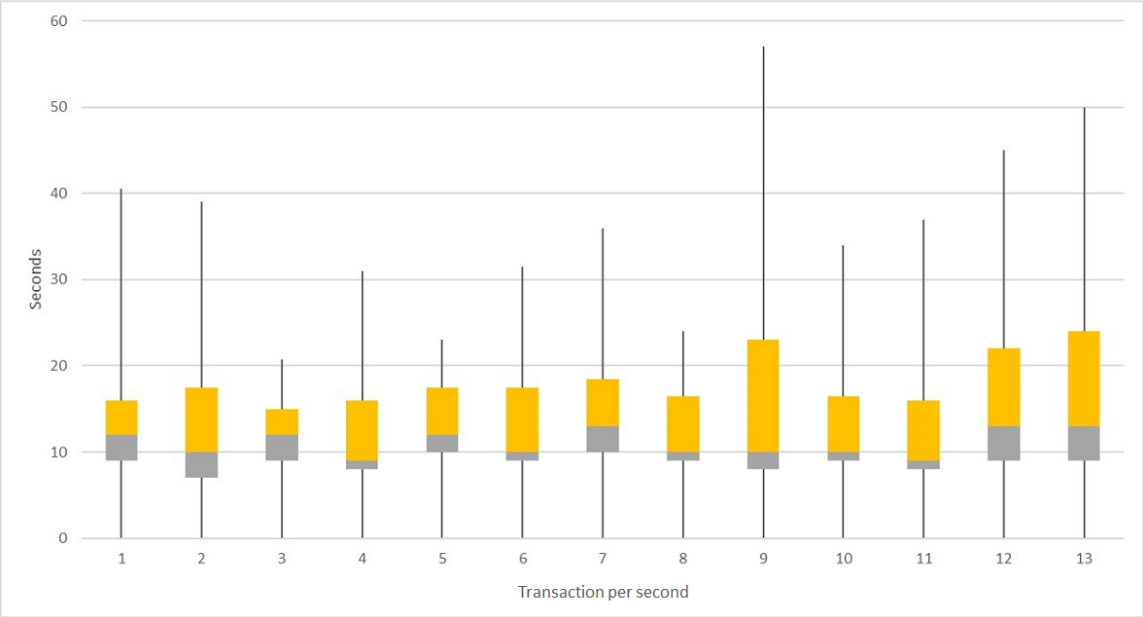

Figure 6. Box plot of simulation results for 1 to 13 transactions per second group. The boxes show the different quantiles of time costs related to different scales of transactions per second.

3.5 Discussions

This chapter described an augmented layered blockchain system in development for most healthcare applications involving data coordination across multiple healthcare facilities. The design of this layered architecture provides generic functions and methods for application developers to securely collect data from different sources without requiring extensive experience in blockchain technology. The layered architecture allows users the ability to audit the legitimacy of previously occurring transactions but prevents users from modifying any components in the blockchain. The features of blockchain provide a solution to current data coordination challenges. The blockchain-based approach extends the ownership of the EHR dataset to each patient. On the basis of decentralized features of

blockchain technology for peer-to-peer transactions, this approach can greatly reduce the healthcare dataset sign-off and release. Data security and authenticity are also guaranteed by the immutability of the blockchain and smart contract-regulated data exchange. Through our simulation process, our system empirically proved the feasibility of the architecture for healthcare applications. In the next chapter, we will have a comprehensive blockchain design to achieve the patient-centric HIE.

CHAPTER FOUR – PATIENT-CENTRIC HEALTH INFORMATION EXCHANGE

We have published this chapter in the Journal of Biomedical and Health Informatics in 2020 [35]. This chapter adopts its main contents with minor modifications.

Zhuang Y, Chen YW, Shae ZY, Shyu CR. Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation. J Med Internet Res. 2020 Jul 27;22(7):e19029. doi: 10.2196/19029. PMID: 32716300; PMCID: PMC7418010.

4.1 Background

Electronic health record (EHR) systems are widely used worldwide [61] with a more than 96% adoption rate among non-federal acute care hospitals in the USA [62]. Timely Health Information Exchange (HIE) across healthcare systems exhibits tremendous benefits in reducing healthcare costs, improving quality of care, and reinforcing disease surveillance [63] and also in improving the quality and efficiency of the clinical research process such as patient recruitment. The Office of the National Coordinator (ONC) for Health Information Technology has spent billions of dollars to achieve meaningful use of EHR and facilitate the development of HIE systems [64]. There has been some success in achieving HIE among business entities such as state-wide hospital systems in the same collaborative association [65, 66]. However, various forms of HIE, listed in Table 1, pose challenges related to data quality [67], data security, patient privacy [6, 68], and patient engagement [69]. In addition, there are recent signs of shifting to patient-centered interoperability [65, 70]. Although one of the three existing HIE forms, consumer-mediated

exchange, allows patients to access and manage their health information online, to achieve a true patient-centric HIE, the patients should have full control of their data. Such controls include authoring healthcare facilities' data access, determining sharable information, acknowledging the data use [71], and approving the life cycle of shared data.

Table 1 Three forms of Health Information Exchange

HIE forms	Definitions
Directed Exchange	Allowing pairs of care providers to share the patients' information used for coordinated care
Query-based Exchange	Giving providers the ability to collect a specific patient's records from among different providers often used for unplanned/ emergency care
Consumer-mediated Exchange	Letting patients control the sharing of their own electronic health information to assist coordinated care and unplanned care.

There are various conceptual models for different HIE forms: (1) centralized model using a central repository to store and manage all patient's health information, (2) federated model consists of a state-wide central HIE patient registry or record locator service (RLS) contains a combination of patients' identifiers to match the patients across multiple regional authorities which maintain the ownership and control over the regional healthcare facilities' records, and (3) hybrid model combines the centralized and federated models using a centralized data repository as the national central authority or RLS to locate patient's records from different healthcare facilities [72, 73]. The existing models have

achieved a certain degree of success of three existing forms of HIE. However, to provide patients a robust and interoperable patient-centric HIE system, the existing HIE models have shown multiple challenges such as security and privacy concerns caused by central repository storage of data or patients identifiers [8], data ownership still controlled by authorities [74], mismatching of patients using RLS [75], and data breach caused by external cyberattacks and the threat of internal fraud [76]. Emerging technologies, such as blockchain, may provide potential solutions for the needed function in the patient-centric HIE system to tackle the aforementioned challenges.

The following scenario illustrates the barriers of the current HIE process to provide patient-centric services: A patient lived in Los Angeles, California between 2000 and 2015 and moved to Columbia, Missouri, where he is a resident currently. He has a medical history of congestive heart failure since 2010 (well managed on medications), and a prior history of alcohol dependency (in continuous remission since 2005). While visiting New York City, he is admitted to an emergency department for shortness of breath. It is critical for the clinicians at the healthcare facility in New York City to access his prior records from providers in Los Angeles, California, and Columbia, Missouri. The patient selected to share only cardiology data but did not want other providers to know his history of substance abuse for reasons of privacy, concerns about provider bias, and recent assurances from his current primary-care physician that his remote history of alcohol dependency has no current relevance for the management of his congestive heart failure. As of today, for the traditional HIE process, this health information exchange will start with a request to the Hartland Regional Health Information Organization (RHIO) and then connect to

Western and Midwest RHIOs to access EHR data from California and Missouri, respectively, through Regional Gateway Connections.

Patient-centric exchange is needed because there are two major barriers for this information exchange to happen: (1) the time required for the RHIOs to locate the patient's prior records without knowing the patient's protected health information (PHI) from the remote healthcare facilities where the patient visited previously [8]; and (2) the vulnerability of the patient's history of substance abuse being accessible to the provider against the patient's will [8, 71, 77]. However, there are three challenges to patient-centric exchange across institutions: (1) security and privacy concerns that may result in appalling financial and legal consequences [78-80]; (2) data breaches caused by unauthorized access of the patients' health records [81]; and (3) data inconsistency between the remote provider's EHR data and the recipient's data [82, 83].

4.2 Rationale Using Blockchain Model for Patient-centric HIE Applications

Based on these barriers and challenges, disruptive technologies such as blockchain may provide feasible solutions by utilizing blockchain features, as shown in Table 2. This work demonstrates the feasibility of applying blockchain for HIE with unique settings using the principle that patients should have ownership of their EHR data to achieve patient-centric HIE. We have also conducted a large-scale patient-centric HIE simulation from granting permission by patients to receiving data by clinicians.

Table 2 Blockchain Solutions for Patient-centric HIE Challenges

HIE Challenges	Blockchain Solutions
1. Difficulty of timely match a patient across different healthcare facilities	Public/ private key pair can be used to represent patients' identities
2. Potential data inconsistency concerns due to integrity loss during transmission	Immutability feature can ensure data consistency
3. Locating healthcare facilities to collect the patient-agreeable information	Smart contract can be utilized to store touchpoints for clinicians to quickly select
4. Potential security and privacy concerns specified as data breaches caused by unauthorized access to health information	"Unhackable" peer-to-peer network ensures every transaction needs patient's authorization

4.3 Methods

To utilize the unique technological capabilities of blockchain for patient-centric HIE, we have implemented a private Ethereum blockchain system with multiple smart-contract functions. A private blockchain is also called a "permissioned blockchain" which limits access to certain users. The system architecture, shown in Figure 7, contains two modules: (1) the *Linkage module*: a system administrator from each healthcare facility will create a touchpoint for each patient's visit after the EHR is ready and input the related primary information into a smart contract for future indexing (as shown in Figure 8), (2) the *Request module*: patients grant clinicians permission to access their data by adding clinicians to the "allowed list" in the smart contract. Clinicians can select records through

the touchpoints after being granted access to the patient's records without identifying the hospitals storing those records. The subsequent exchange of data among the involved remote healthcare facilities will include data encryption and the use of the blockchain system to send and retrieve decryption keys (as shown in Figure 9).

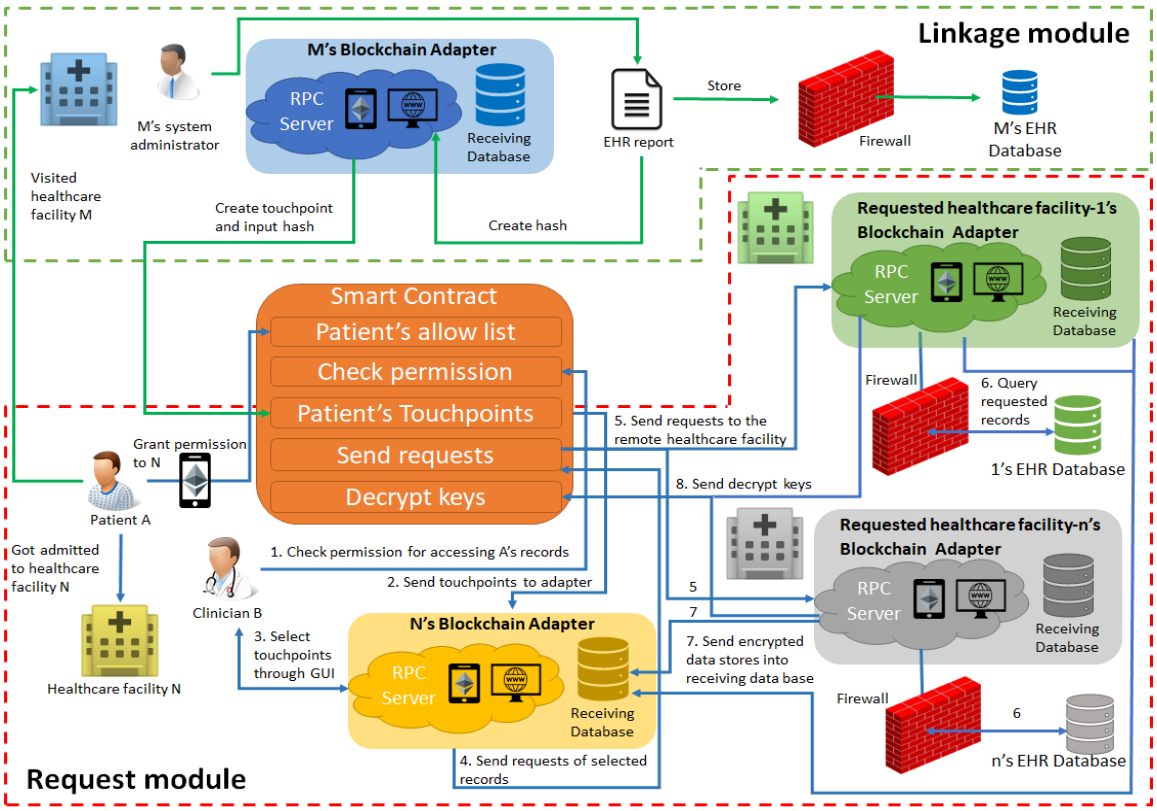

Figure 7. System architecture with two modules (the Linkage module links the EHR databases with the blockchain by creating touchpoints to index the records in the future; the Request module allows patients to give permission to clinicians to access their data through blockchain and to request records by selecting touchpoints through the blockchain adapter).

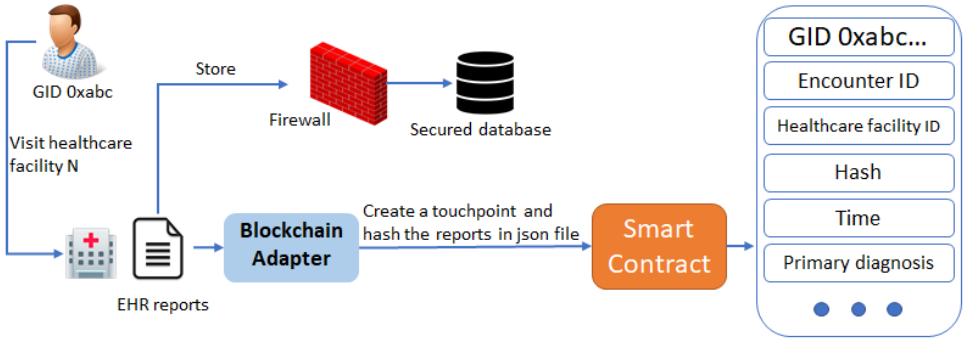

Figure 8. The blockchain adapter extracts metadata and hashes the EHR reports in JSON format, stores this information in a smart contract, and stores the EHR data in the secure database.

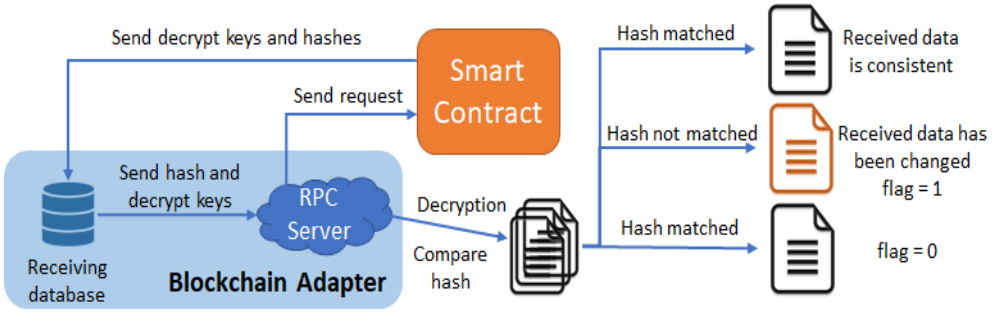

Figure 2. The blockchain adapter retrieves decryption keys and hashes from a smart contract to decrypt the received EHR data, then hash the data using the preinstalled hashing function and compare it with the original hash; any mismatched records will be marked.

4.3.1 Linkage Module

When the EHR data is ready for a patient's visit, the healthcare facility's adapter will hash the entire visit record in a JSON file and store the hashing value in the smart contract along with the touchpoint before the EHR data is stored in the secure database. The hashing value will be used for verifying data consistency in the data decryption step. Any modification of the data, even initiated from the healthcare facility adapter, intentionally or unintentionally, will result in unmatched hashes and security alerts after final decryption (Challenge 2 of Table 2).

Once the smart contract is deployed into the blockchain, the blockchain returns a smart-contract address and an ABI; this, rather than the smart-contract code or the data stored inside the smart contract, is viewable by all users. Using the smart contract to store the touchpoints can keep the touchpoints secure, immutable, anonymous, and easily searched by the patients and the authenticated clinicians. Figure 10 shows the source code for the inputting touchpoint function and its ABI in the smart contract.

```
function input_touchpoints(
                        string _GID,
                        string _Encounter_ID,
                        string _facility_ID,
                        string _hash,
                        string _time,
                        string _PD
                        ) public {
    records memory record;
    record.GID=_GID;
    record.Encounter_ID=_Encounter_ID;
    record.facility_ID=_facility_ID;
    record.hash=_hash;
    record.time=_time;
    record.PD=_PD;
    patient_EHR[GID[_GID]].push(record);
}
```

```
{"constant": false,
    "inputs": [
        {"name": "_GID",
            "type": "string"},
        {"name": "_Encounter_ID",
            "type": "string"},
        {"name": "_facility_ID",
            "type": "string"},
        {"name": "_hash",
            "type": "string"},
        {"name": "_time",
            "type": "string"},
        {"name": "_PD",
            "type": "string"}
    ],
    "name": "input_touchpoints",
    "outputs": [],
    "payable": false,
    "type": "function",
    "stateMutability": "nonpayable"}
```

Figure 10. The source code of inputting touchpoint function is coded in Solidity which shows on the left; the ABI of this function, which only contains structure information, is shown on the right.

4.3.2 Request Module

After a patient is admitted to a healthcare facility, it is unrealistic for patients to authorize each of the clinicians in some situations, such as an emergency room visit when many clinicians are involved during the patient's care. The healthcare facility will be assigned an umbrella account in the blockchain that links to all clinicians involved in the care. All the clinicians could access the patient's records with one-time authentication (adding the shared account into the "allowed list") from the patient. The access history will be recorded to the blockchain and the auditing of individual clinician's access to the

patient's record will be managed by the local access control within the healthcare facility. The patient can add the facility's umbrella ID to the "allowed list" through biometric authentication or a web-based GUI, as shown in Figure 11. The clinician's proxy ID should be automatically populated into the GUI after the patient and the clinician provides biometric information to authenticate the system. Only the clinicians under this umbrella ID can access the patient's data through the GUI for clinicians, as shown in Figure 12.

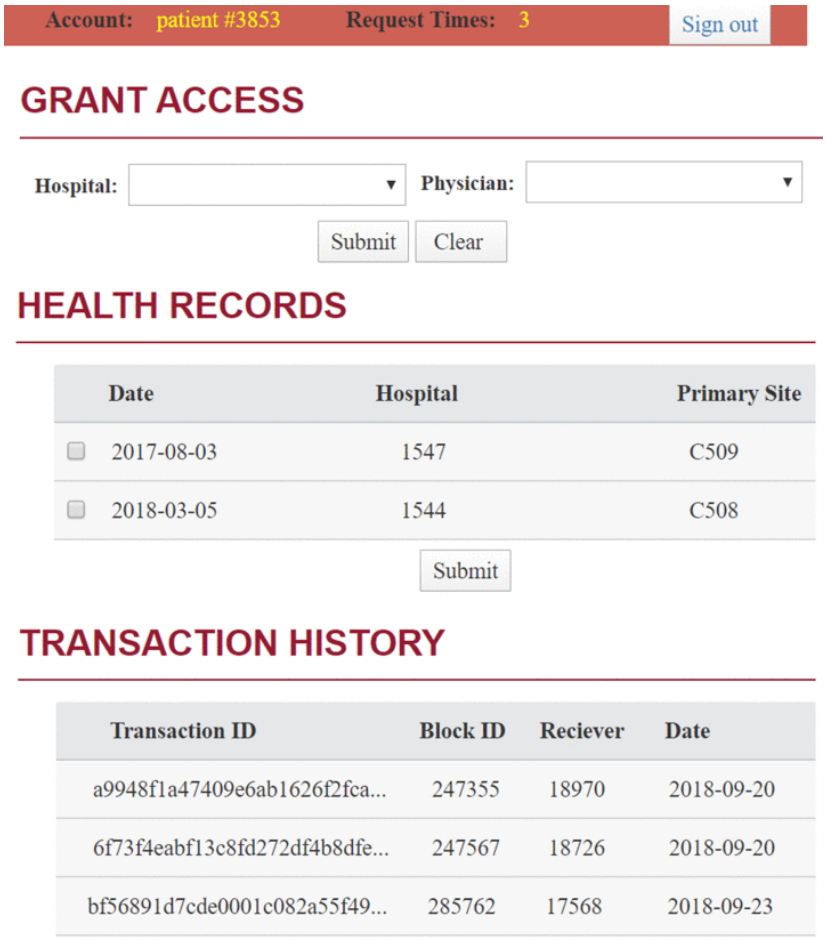

Figure 11. GUI for patients to grant clinician permission, check personal EHR reports, and check who has accessed their records in the past (showing the transaction ID of the accessing clinician ID, and the date); patients can personalize the data segmentation after retrieving their health records.

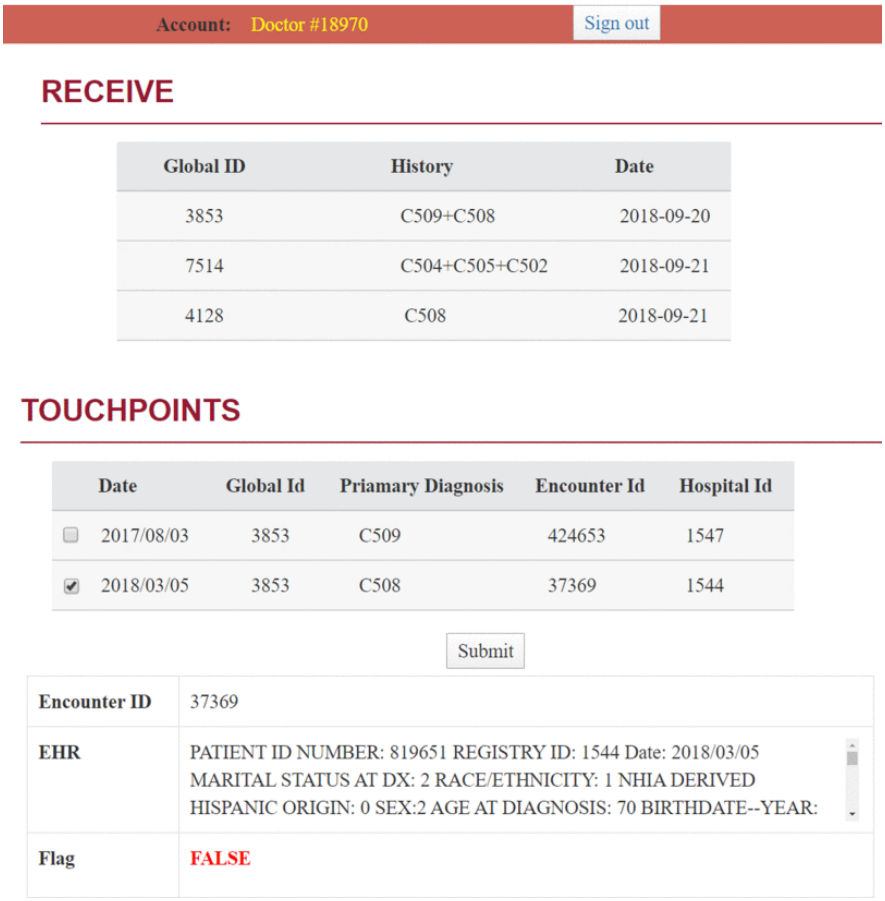

Figure 12. GUI for clinicians to check received patient records, showing a summary of EHR record for patient #3853 after selecting the exact visit; the flag shows "False" because the hash values don't match since we have intentionally modified the data.

After the patient's consent is recorded in the blockchain, the HIE process takes the following steps (as enumerated in Figure 7): (1) the clinician confirms that his/her clinician ID has been added to the patient's allowed list; (2) the clinician receives the touchpoint list from the smart contract; (3) the clinician selects the touchpoints related to this visit through the GUI; (4) the touchpoint selections are sent to the smart contract through the healthcare facility's blockchain adapter; (5) the smart contract uses the recipient's healthcare facility's adapter to request the selected records chosen from the remote healthcare facilities' adapters; (6) the remote healthcare facilities query the records inside their EHR systems; (7) all selected EHR data is dynamically encrypted by the remote healthcare facilities'

adapters and temporarily stored outside the clinician's facility's firewall; the encrypted data's locations and hash values are sent to the receiving database in the adapter of the clinician's facility; and (8) the remote healthcare facilities' adapters will automatically encrypt the decryption keys with the recipient's public key and send the encrypted decryption key to the smart contract. The recipients' private keys will be stored at the home hospital's adapters and protected by the hospital's security policy.

Selecting touchpoints saved into blockchain from each visit, the clinician can quickly check the records related to the visit instead of browsing all the historical data. This function offers efficient information retrieval for the clinician in order to have a better sense of the patient's medical history (Challenge 3 of Table 2).

Two layers of data security are implemented, at both the smart contract level and user application level, to the original hashing algorithm of the Ethereum blockchain. The smart contract level defines multiple modifiers on smart contract functions, meaning that only selected roles can execute certain functions. For example, after the recipient's blockchain adapter automatically retrieves the decryption keys from the smart contract and the encrypted data, only the clinicians in the "allowed list" can retrieve the decryption keys. The decryption process will then automatically run on the recipient's healthcare facility's adapter using the clinician's private key to decrypt the decryption keys. This process then decrypts the data using a predefined encryption algorithm and a preinstalled hashing algorithm to hash each decrypted visit. The adapter automatically compares the hashing value with the original hash that has been retrieved from the touchpoint. Any modification of the data from the original source, or in the transition, will result in a mismatch of the hash and the record will send an alert in red font. For example, as shown in Figure 12, we

falsified patient #3853's single record in the database (by changing the patient's value for the race from “2” to “1”) after the touchpoint and hash were stored in the smart contract, which resulted in a flag showing “False” even when the encrypted data was decrypted.

In our scenario, the patient does not want to show his history of substance abuse and can choose to hide this information from the touchpoints, but the hidden records will not be removed from their records. Patients can always recover the original list after data segmentation. After the decryption process, the recipient's healthcare facility's adapter will hide the information from the decrypted data and will not show it to the clinician.

The user application level is defined by the smart contract based on each role, such as the future access mechanism for clinicians. While intuitively the remote site could choose to store all exchanged data**,** in our design, a policy is required for each healthcare facility's blockchain adapter to either delete, partially keep, or set a life cycle of the shared data in the local facility based on patients’ permission. This mechanism ensures that all of the exchanged information should be only used with the patient's consent which could be granted for future use (permanently stored in the EHR of the remote facility) or one-time use (immediately revoked after care is completed and updates are sent back to the home facility). Furthermore, we have set up a trigger for the local databases in all the adapters; once the encrypted data is queried from the database for decryption, the decryption key and encrypted data will be deleted from the database to prevent future data access without patients’ consent and to clear storage space for future transactions. The smart contract will monitor the process in each blockchain adapter to enforce the policy, in order to minimize the data breach problem (Challenge 4 of Table 2). Patients can also revoke permission

through the GUI if they have mistakenly input a clinician's ID or an umbrella healthcare facility ID by removing the ID from the “allowed” list.

4.4 Simulation

We set up five computing nodes representing five different healthcare facilities. Each node was installed on an Ubuntu 16.04 system and Apache HTTP Server. The starting node initialized the blockchain and the smart contract was deployed. The other four nodes joined the blockchain by going through the setup procedure described previously. We created 20,000 patient accounts in total and 100 clinician accounts for each healthcare facility node. The simulation randomly selects patients to grant clinician access data for multiple healthcare facilities based on patient preferences.

We used the Surveillance, Epidemiology, and End Results (SEER) dataset [46] for the simulation. We selected 80,000 records with 133 attributes from the original dataset and generated a PHI (date of birth and email address), visit date, healthcare facility ID and patient ID to be added to each record. These data were distributed and stored in four nodes depending on the healthcare facility ID. We created these scripts to simulate the whole HIE process: (1) load the touchpoints into the blockchain from different adapters depending on the visit location; (2) randomly select five patients to respectively grant access to five different clinicians from different healthcare facilities; (3) randomly select several patient's records from the touchpoints list and request these records by authorized clinicians from their healthcare facilities’ adapters; (4) query the requested records by each selected remote healthcare facility's adapters; and (5) encrypt all the queried records in the adapters and send the encrypted data to the smart contract locations as URL pointers and decryption keys. Scripts #2 and #3 were run every five seconds to balance the memory load for running

scripts on each blockchain node. Script #4 was run twice per second to detect the requests. These steps not only simulate the process of patient-centric HIE but also test the stability of the system.

The adapters at the simulated recipients' healthcare facilities retrieved the data locations and decryption keys from the smart contract, and another script decrypted the data and hashed each record to compare with the original hashes; but because these steps were performed off the blockchain, statistical summaries were not completed for them. We manually falsified several records to test the data integrity function (as shown in Fig. 4).

In order to test the robustness of the system, we also randomly stopped nodes during the simulation. The other nodes continued to work, but all the requests sent to the stopped node's adapter could not be executed. The result was that the affected transactions were still approved and stored in the blockchain, but the recipients' healthcare facility could not receive data from the remote healthcare facility simulated by the stopped node. After restoring the node to service, the adapter automatically found the previous peers and synchronized itself with the blockchain for the missing period within seconds. The blockchain will only stop working if all of its supporting nodes stop working simultaneously.

4.5 Results

We simulated 1,553,635 data request transactions in four months by running the scripts continuously. One hundred percent of the transactions were successfully approved, and their encrypted queried data was stored in the requesting facilities' databases. The box plots in Fig. 7 report the total processing times for clinicians to receive (1) permissions after being added to the "allowed list" by patients and (2) the decryption keys provided by

different remote healthcare facilities involved in the HIE process. If a clinician requests the patient's records from all four other healthcare facilities, the clinician would receive four separate decryption keys sent by all healthcare facilities. Table III lists the statistics of the processing time of HIE procedures after patients grant clinicians permission to access the records. On average, clinicians received permissions as well as the metadata lists in 20.398s and retrieved the encrypted data's locations with their decryption keys in 23.844s.

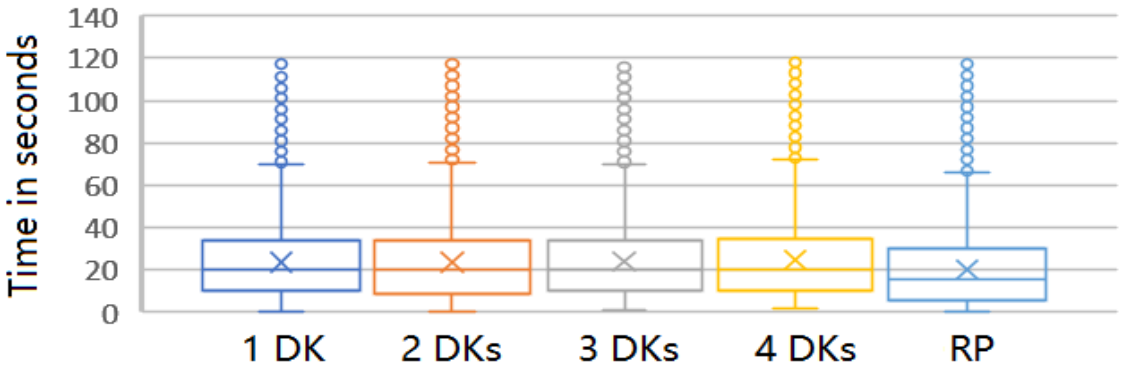

Figure 13. Box plots of time for clinicians to receive permission (RP) and time to receive decryption keys (DK) from different facilities.

Figure 14. Time to generate new blocks containing different numbers of transactions.

Table 3. Results of Processing Times of HIE Procedures

	Request #	Mean	Min	Max	Stdev
Receive permission	1,533,620	20.398 s	< 0.01s	122s	22.217
Receive encrypted data, decryption key	1,533,620	23.844 s	< 0.01s	121s	19.2800

Most transactions were validated and written into a block in about 23 seconds. Receiving decryption keys from a different number of nodes did not significantly affect the receiving time. The access-granting process took insignificantly less time than retrieving decryption keys. Fig. 8 shows the distribution of time elapsed to generate blocks with different transaction volumes; most blocks required around 40 seconds. The maximum number of transactions occurring in one block was 274, with a validation time of 78.37 seconds. The longest validation time is 122.54 seconds with 77 transactions occurring in the block. Block generation times in our experiment were varying but reasonably stable.

4.6 Discussions

The unique contributions of this work include providing the following practical characteristics to the blockchain system to achieve patient-centric HIE: (1) blockchain adapter setup to communicate with blockchain, process the sending/receiving healthcare records, and provide graphical user interfaces for users to have a better visualization of the interaction with the blockchain system, and (2) two layers of security settings to ensure that only authorized users can execute certain smart contract functions and minimize the data breach problem, and (3) a hashing mechanism to ensure data consistency and (4) personalized data segmentation gives patients the ability to control of their records by choosing only the information they would like to share, and (5)

touchpoint selection for clinicians to select the health records that related to the visit without browsing through entire records, and (6) a large-scale simulation using the implemented proposed model to evaluate the feasibility, stability, and robustness of the proposed blockchain model for the HIE application.

It is noteworthy to mention that blockchain technology is not the only solution for HIE. This chapter demonstrates the feasibility and robustness of using the unique features of blockchain technology in HIE for the health IT community to consider applying the variations of the blockchain technology for HIE tasks, as well as to evaluate regulations and policies to adopt this emerging technology.

CHAPTER FIVE – PATIENT RECRUITMENT

We have published this chapter in the AMIA Annual Symposium proceedings in 2019 [84]. This chapter adopts its main contents with minor modifications.

Zhuang Y, Sheets LR, Shae Z, Chen YW, Tsai JJP, Shyu CR. Applying Blockchain Technology to Enhance Clinical Trial Recruitment. AMIA Annu Symp Proc. 2020 Mar 4;2019:1276-1285. PMID: 32308925; PMCID: PMC7153067.

5.1 Background

Patient recruitment is essential to the success of clinical trials. Failure to meet recruitment goals in time results in a waste of funds and time, incomprehensible statistical results, and delay of the study timeline that could double the planned recruitment period [85, 86]. 86% of clinical trials don't achieve their recruitment goals on time [87] and 19% of registered clinical trials were either closed or terminated due to failure to reach expected enrollment [88]. Barriers persist although there have been many research papers addressing the challenges of identifying and recruiting subjects to clinical trials over the past decades [85, 89-92].

Barriers to recruiting patients into clinical trials can be classified into three different categories (Table 4) based on (1) sponsor perspectives, (2) principal investigator perspectives, and (3) subject perspectives [2, 92]. Sponsors initially need adequate participants for the potential trial to file an application with the Food and Drug Administration (FDA) for approval [2]. Inefficient advertising models such as radio, newspaper, physician referrals, flyers, cold calls, etc. make it difficult to meet the expectations of initial recruitment on time [93]. Sponsors need to design clinical trial

protocols such as inclusion/exclusion criteria which can be used to check the eligibility of potential subjects. Detailed protocols can drastically narrow the subject population, which increases the difficulty of recruitment [2, 92, 94]. Sponsors need to settle on trial sites without knowing the geographical distribution of future subjects [91], but distant trial sites will deter many potential subjects [92]. For principal investigators, barriers include lack of awareness of available, appropriate clinical trials, excessive time spent to get the informed consent of participants, and insufficient trial protocols [88, 91, 94]. Subject-related barriers are related to participation, such as patients' lack of awareness of the available clinical trials; difficulty understanding complex protocols; high expenses if the trial has no clinical sites nearby; and distrust in the clinical trials [2, 92, 94, 95]. With these persistent challenges, an efficient model is needed to enhance the recruitment process.

Table 4. Current recruitment barriers from different perspectives

Sponsors	Principle investigators	Subjects
1. Inefficient advertising models 2. Protocol limitation 3. Beforehand trial sites selection	1. Lack of awareness 2. Time consumption 3. Insufficient trial protocols	1. Lack of awareness 2. Complex protocol 3. Inaccessible clinical sites 4. Distrust of clinical trials

5.2 Methods

To utilize the unique technological capability of blockchain for clinical trial recruitment, we implemented a private Ethereum blockchain system to simulate the recruitment process. The authority's node as the creator node needs to start the blockchain system using a unique "genesis block" file, and other nodes and adapters then join the

system using the blockchain identifier and IP address of the authority's node. The whole system architecture (Figure 15) contains two modules: (1) A master smart contract is used for auto-matching of potential subjects for all trials using inclusion and exclusion criteria as shown on the left of the figure; (2) Multiple trial-based smart contracts are used for patients' enrollments, trial management and future persistent monitoring for different clinical trials as shown on the right of the figure. In this setting, all the users can access the master smart contract so that it can reach any user in the system to perform the matching process. The trial-based contract is only available to the users to participate in that trial so that the trial-based contract cannot notice the users outside that trial. A use case is that the CRC can monitor the subject's condition during the clinical trial through the trial-based contract, but they can't access subjects' records for other clinical trials in order to know who has participated in other clinical trials.

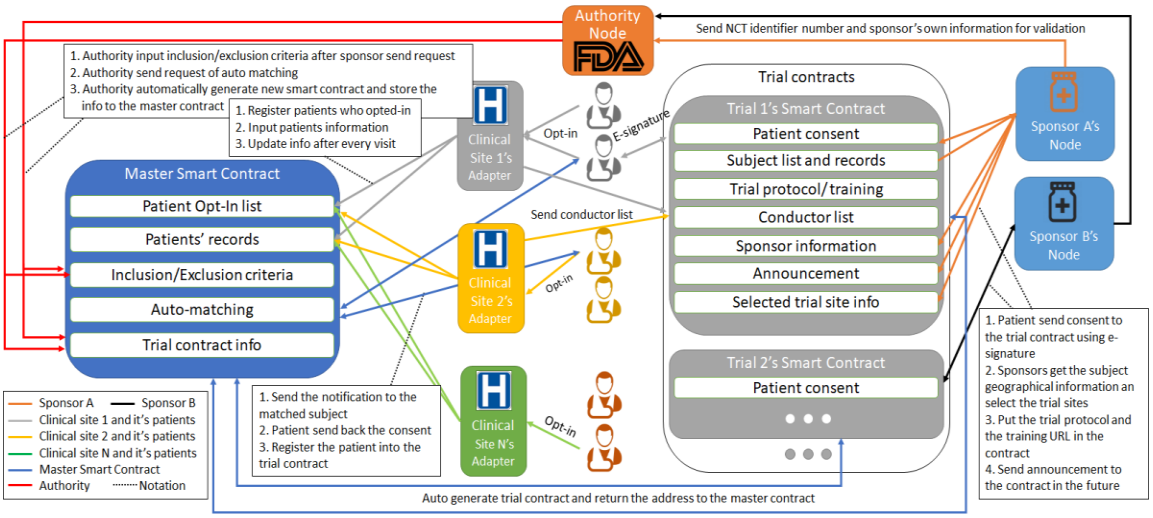

Figure 15. System architecture master smart contract and different clinical trial smart contracts sponsor's account

Only the selected group of users can execute specific functions in smart contracts. Users without privileges cannot see the data stored inside the smart contract (Table 5). To

ensure the accuracy of input data, inclusion/exclusion criteria need to be input by the sponsors and executed by the authority, and the patients' primary records which include demographic information, previous primary diagnosis, and treatment from each visit that used for trial matching can only be input by the clinical sites. The authority has an oversight role in the system. All the clinical sites and sponsors need to get approval from the authority to provide a node to join the system. The authority will intervene in any inconsistent data such as differing patient records in the master smart contract and trial-based contract. The authority can trace the inputter of the records and investigate the reason. This setting can ensure the trial is conducted precisely under the authority's real-time surveillance.

Table 5. Privileges for users to use smart contract functions

		Execute	**View**
Master Smart Contract	Patient Opt-in list	Clinical sites	Clinical sites/ Authority/ Opted-in subjects/ Sponsor
	Patients' Information	Clinical sites	Clinical sites/ Specific patient
	Recruitment Criteria	Authority	authority/ Specific sponsor
	Auto-matching	Clinical sites	N/A
	Trial contract info	Authority	Trial sites/ Authority/ Enrolled subjects/ Trial sponsor
Trial Contract	Sponsor information	Authority	All
	Patient's trial records	CRC	Trial sites/ Authority/ Specific subjects/ Sponsor
	Other functions	Sponsor	All

Module 1 requires all clinical sites to input the opted-in patient list and their primary records. Each sponsor needs to send a transaction to the authority containing the NCT identifier number and its own information for validation. After the authority validates the sponsor's identity and the authenticity of the requested trial, the authority will input the inclusion/exclusion criteria to the master smart contract. In the meantime, a trial-based contract for this clinical trial will be generated and the contract address will be stored in the master smart contract. The sponsor can request the auto-matching process after sending the transaction to the authority. Part of the auto-matching smart contract code and the returned ABI after compiling this function is shown in Figure 16.

(a)

```
//sample of matching function
function auto_inclusion_matching
//input criteria through GUI
(uint _agemax, uint _agemin, uint _gender, uint _diagnosis)
//only sponsor can execute this function
public onlySponsor{
    //traverse the patients' records
    for(uint i=0 ; i < patients.length; i++){
        //patient_info is the structure
        //check each criteria with previous records
        if (patient_info[patients[i]].age > _agemin){
            if(patient_info[patients[i]].age < _agemax){
                if(patient_info[patients[i]].gender == _gender){
                    if(patient_info[patients[i]].diagnosis == _diagnosis){
                        matched_inclusion.push(patients[i]);
                    }
                }
            }
        }
    }
}
```

(b)


```
{
    "constant": false,
    "inputs": [
        {
            "name": "_agemax",
            "type": "uint256"
        },
        {
            "name": "_agemin",
            "type": "uint256"
        },
        {
            "name": "_gender",
            "type": "uint256"
        },
        {
            "name": "_diagnosis",
            "type": "uint256"
        }
    ],
    "name": "auto_inclusion_matching",
    "outputs": [],
    "payable": false,
    "stateMutability": "nonpayable",
    "type": "function"
},
```


Figure 16. Sample matching function and ABI ((a) The sample code of the matching function. The real function is calling several matching functions based on the criteria. (b) The ABI of the sample function which is viewable to every user.)

Patient matching is a two-step process. The first step is using the auto-matching function which can only narrow down the patient selection but cannot perform precise matching due to the complexity of the inclusion/exclusion criteria. The master smart contract will automatically match the criteria with the patients' records stored in the smart contract to select potential subjects. Referring to the example shown in Figure 10, the master smart contract filtered out patient B due to the exclusion criteria of history of renal

disease. The second step is precisely matching by the clinical sites through checking the auto-matched subjects' EHR in the hospital's secured database with the patients' consent and sending the result back to the blockchain through its own blockchain adapter. The exclusion criteria in the example in Figure 17 also have current tobacco use which is not recorded in patient A's information. The master smart contract will notify patient A that there is a potential clinical trial that he/she might be eligible to participate in and need his/her authentication for the sponsor to access his/her EHR to double-check with the details. The sponsor can communicate with the clinical sites which patient A visited before to check whether he/she uses tobacco currently with patient A's consent by E-signature using the private key. The clinical site will perform precise checking for the sponsor.

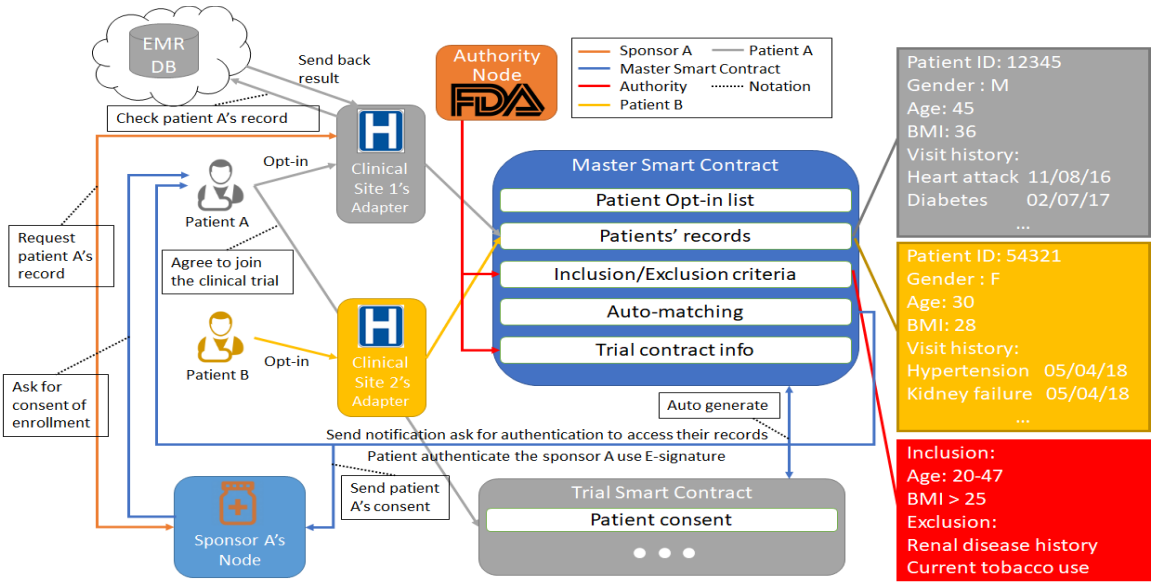

Figure 17. An example of a matching process using smart contract and ask consent to join the clinical trial.

Module 2 is about how subjects send consent to join the trial chain. If patient A has fully matched the criteria, sponsor A will send a transaction to patient A to ask for enrollment, patient A can agree to join the trial using an E-signature. Then all of patient A's primary records will be stored in the trial-based contract through a private transaction

which means the records are only visible to the smart contract. Sponsor A can select the trial sites based on the density distribution after collecting all the enrolled subjects' geographical information. The patients still need to go to the clinical sites to sign all paperwork and proceed with the trial. The sponsor can use the trial-based contract which can only be accessed by the participants of this trial publishing the trial detail and announcement.

We have also built a web-based GUI on the RPC servers for users to better interact with the system instead of using plain command codes in the blockchain console. To log in to the GUI, all users need to set up a username and a password. The username and password will be stored in the local RPC and mapped to the public key and private key. This could potentially be implemented using personal biometric information to log in to the system. All the functions showed in the GUIs are sending or receiving data from the blockchain rather than a cloud database. Different roles of users will have different GUIs to operate the system. The clinical site's GUI will have the functions as input the patient's primary visit records to the smart contract; check ongoing clinical trial as a trial site; check requests from sponsor to check on specific patient's eligibility and send the result back to the blockchain (Figure 18(a)). The potential matched clinical trials requesting authentication will be displayed in the patient's GUI. After the patient clicks "approve", a transaction signed by the patient's private key will be sent to empower the sponsor to request precise matching from the clinical sites. Patients can also check their basic information and the visiting records but cannot be changed. Patients can also check enrolled clinical trial information the same as the trial sites (Figure 18(b)). The sponsor's GUI will have a list of ongoing, recruiting and completed clinical trials. For the recruiting

clinical trials, they will show the list of matched subjects. Once the request is sent, it will deliver a transaction to the patient through the sponsor's account requesting authentication (Figure 18(c)). The authority will have all the clinical trial lists and the trial details (Figure 18(d)).

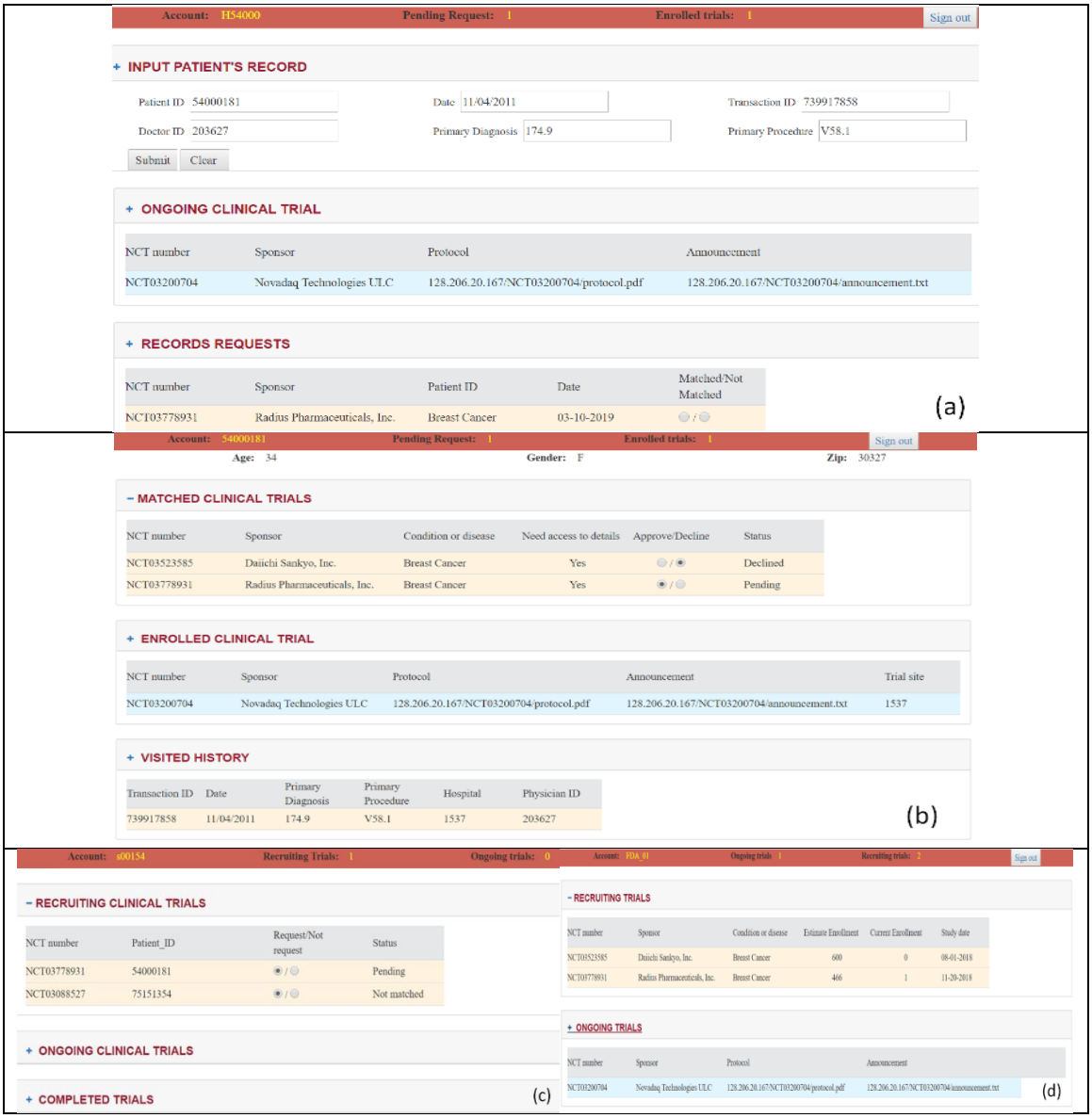

Figure 18. GUIs for (a) clinical sites to input primary records and receive requests from sponsors, (b) patients to receive notifications and authenticate sponsors, (c) sponsors to request a precise match for potential subjects, (d) authority to monitor all the trials

5.3 Simulation

To test the feasibility and efficiency of the system, we simulated the recruitment process which started from the moment that the authority registered the trial criteria to the master smart contract, and continuing to the end of the process when recruited subjects were added to the trial-based contract. We have randomly picked 10 currently recruiting clinical trials. Criteria such as specific medication use vary from different clinical trials. It is doable to code all the criteria into the master smart contract. To demonstrate the feasibility of using the system for recruitment, we only select the frequent criteria among those 10 trials for the simulation as listed in Table 6. We haven't simulated the precisely matching process since it is an off-chain process that needs clinical sites to manually check whether the auto-matched patients' EHR has fully matched the criteria. All the frequent criteria we have selected can be matched from the auto-matching function directly. Five blockchain adapters are set up using the Intel NUC machines: (1) Authority (simulated authority) node as the starting node, (2) sponsor node, and (3) three different clinical sites. We have set up 2,000 synthesized patient accounts on each clinical site's node. We use the data from the SEER program for the simulated cases [96]. For our simulation, we randomly picked 6,000 patients with breast cancer from the SEER dataset and evenly distributed them into three clinical sites.

Table 6. Frequent criteria for breast cancer clinical trials among ten selected recruiting trials with frequency

Inclusion criteria	1. 18 years and older (9 times) 2. Female (6 times)

	3. Diagnosed as breast cancer (7 times) 4. Negative metastatic involvement (3 times)
Exclusion criteria	1. Stage IV cancer (8 times) 2. Pregnant or breastfeeding (8 times) 3. Persistent malignant (4 times) 4. breast implants (3 times)

We have created one database on each clinical site's adapter. Before the simulation process, we have written scripts to populate all the patients' previous primary diagnoses and treatments into the master smart contract through each clinical site's administrator's account. We have manually input the inclusion/exclusion criteria to the master smart contract through the authority's account that comes with a trial-based contract generated automatically. The returned ABI and address will be stored in the master smart contract associated with the NCT number. After the sponsor executes the matching function, it will take several seconds to run the function and for users to validate and write into the coming block. Matched subjects will receive a notification asking for enrollment. We have randomly selected a list of patients to agree to join the trial, then all of their information such as geographical location, patient ID, and demographic data will be sent to the sponsor. The sponsor can select the appropriate trial sites after receiving the enrolled patient's geographical distribution. Assuming all the matched patients have agreed to join the trial, we have written a script to register all the patients to the trial-based contract through the sponsor's adapter.

5.4 Results

Due to the limited information of the SEER database, we cannot check the pregnancy in the exclusion criteria or whether the breast cancer is persistent malignant. After executing the auto-matching function in the master smart contract, 1,145 patients out of 6,000 patients are matched in 2.13 seconds. We have used SQL to query the databases on the same criteria and got the same results. We have randomly selected 100 matched patients to join the trial smart contract using the trial address stored in the master smart contract by sending the consensus transaction from their patient account. We created a control script in each adapter to only send five transactions from different patient IDs to the blockchain every second. All the patients have been successfully registered in the trial-based contract without breaking the chain. Figure 5 shows the results of calling the master smart contract and trial-based contract's functions through the trial sponsor's account as: (1) checking the total matched patients for the trial NCT 103200704; (2) getting the trial-based contract address (the ABIs of trial-based contracts are the same and pre-stored in the adapters for deployment); (3) checking whether the trial-based contract address got matched with the address stored in the master contract; (4) receiving the amount of enrollment for this trial; (5) obtaining the geographical information of all enrolled subjects (only showing part of the 100 enrolled subject's zip codes) for the trial NCT 103200704. The results are from the blockchain console that shows the response of calling smart contracts functions from plain codes. Users will use the GUIs (Figure 11)

rather than the blockchain console to send requests and receive results in practical use. The NCT103200704_contract which is a trial-based contract shown in Figure 12 is automatically generated by the master smart contract.

5.5 Discussions

The blockchain features are a good fit for the clinical trial recruitment process. From the simulation we have done, all the users can see all the recruiting trials' information. Sponsors and clinical sites need to get the authority's approval to join the blockchain system and all the patients' identities are verified by the clinical sites. The transactions are public auditable and also under the surveillance of the authority. The data component in the transactions is encrypted and can only be decrypted by a certain group of users. These features ensure the authenticity of the clinical trial, data security of the transactions, and the accuracy of data exchange that has occurred during the clinical trial. After integrating with the smart contract functions, the blockchain system becomes more feasible for recruitment. The auto-matching function is expected to provide the patients an efficient tool to search potential clinical trials. Only auto-matched patients will get notifications from the sponsors. The patients can comprehend the details of the clinical trial after receiving the notification. The auto-matching process also saves time for the patients to understand the complex recruitment protocol. This process shows how blockchain solves the issue of lack of awareness of the eligible clinical trials. Since the criteria are inputted by the authority and all the clinical sites share the same version of the protocol, there won't be an insufficient trial protocol.

Figure 3. Calling smart contract functions to check the trial NCT103200704's info from the sponsor's account

Using the blockchain system, the patients only need to opt-in and wait for notifications of the matched potential eligible clinical trials.

The trial-based contract mechanism optimizes clinical trial management. Only the trial participants can access the data in the trial-based contracts. Each trial-based contract is isolated from the other. From the simulation results, we have narrowed down the patients' selection pool. Selected patients have been successfully added to the trial-based contract after sending their consent using a private key. The transaction processing time depends on the block generating rate which is defined in the genesis block file. In our system, a new block will be generated around every 30 seconds. The time consumption is acceptable for the subject matching process, solving the issue for clinical sites matching with potential subjects.

The sponsor has received the subjects' list after all the subjects have been registered to the trial-based contract. Then the sponsor can get the subjects' geographical distribution by simply calling the smart contract function. Then the sponsor can select trial sites after considering the subjects' geographical distribution. This approach could potentially reduce some opportunities for clinical sites to join the clinical trials but provides benefits for patients to access the clinical trial easier and sponsors to set efficient sites for the recruitment.

Our design also provides potential solutions to the current challenges for healthcare applications involving the blockchain. The username and password setting is a potential solution to the key management vulnerability. The user can also contact the authority to rebuild a new account and remove the original one. To empirically prove that only legitimate trials can be accessed and matched with qualified patients, we have intentionally

tested malicious behaviors, such as manually changing patient's record to meet a trial criterion, registering a fake clinical trial that does not exist in the authority's database through a sponsor's account, and executing auto-matching process through a clinical site's node, we found that all of these transactions were rejected by the smart contract as expected. These experiments ensure that only transactions that follow the rules listed in smart contracts will be executed.

The user can audit all the transactions theoretically through the blockchain console. However, there are three reasons that there will not be any loss of privacy: (1) the data contained in the transactions sent to the smart contract cannot be decrypted by users; (2) users can only see that all the transactions are sent to different blockchain addresses but cannot know the receivers' identities or decrypt the data contained in the transactions; and (3) users can only call the smart function to retrieve their own data stored in the smart contract; other actions will be discarded automatically.

To demonstrate the scalability of our implementation, we have simulated a clinical trial recruitment process which contains one request from a sponsor with a 2.13 second transaction time and 100 consensus transactions, as discussed previously in the Results section, from 100 patient accounts with a 24.69 second total transaction time. Using the latest report of global participation in clinical trials by the FDA, there were 131,749 total participants during 2015-2016. Our simulation result with a controlled input transaction frequency for five transactions per second through the RPC node leads us to project that it would take several hours for both matching requests and sending the consensus for the entire year. It is noteworthy to mention that our system setting could avoid Ethereum's

scalability limitation by spacing the transactions, but it is not to solve the known scalability issue of Ethereum.

CHAPTER SIX – CLINICAL TRIAL MANAGEMENT SYSTEM

6.1 Background

Because conducting clinical trials involves complex processes, good management is critical to success [13]. The Clinical Trial Management System (CTMS) is a set of software tools used for managing clinical trial processes including but not limited to protocol development, site selections, patient recruitment, study conduct, data collection, data analysis, and study closeout. With the increasing adoption of the CTMS, many substantial benefits such as accessing up-to-date information, improving data quality, and boosting overall study efficiency have simplified the traditional labor-intensive management process [10, 97, 98]. A complete CTMS design must be secure, cost-efficient, compliant with regulations, traceable, and auditable to manage the process for each phase of the study [10-12]. However, the current CTMS market is fragmented and lacks thorough designs with all needed features and management tools [12, 13]. According to the 2019 Unified Clinical Operations Survey provided by Veeva (a global life-science service), nearly all respondents (99%) had issues with their current CTMS and 90% of the respondents reported a significant deficiency on at least one CTMS application [14]. Emerging technologies, such as blockchain, are believed to potentially reengineer CTMS and provide a comprehensive solution [99] since it has unique features such as immutability to ensure data consistency, a peer-to-peer system with public auditability (all blockchain transactions can be audited by any user at any time) to provide regulatory compliance, anonymity (all users are represented by a unique hash string) to protect patient privacy, and a smart contract which is a self-executing programable computer protocol that can be

designed for different applications. Quorum blockchain, a private blockchain developed by J.P. Morgan that requires participating users to gain permissions from the blockchain initiator before joining, has enhanced security, scalability, and efficiency based on the original blockchain [100, 101]. The performance of the Quorum blockchain in areas such as transaction throughput and transaction latency have been evaluated as extraordinarily improved (compared to the original blockchain) using the Raft consensus mechanism for the validation process without compromising its unique properties [100].

We have implemented a blockchain platform that provides the unprecedented software designs for key components of CTMS to achieve better management and monitoring of clinical trials with the following applications: 1) an auditable, sharable, and transparent electronic Trial Master File (eTMF); 2) a fast patient recruitment model with an automated matching mechanism through the smart contract; 3) simplified enrollment by using a digital signature validated by blockchain; 4) a timely Electronic Data Capture (EDC) system that ensures data consistency, traceability, and security through blockchain's properties; 5) a reproducible data analytics module that keeps records of data and code usage; and, 6) a secure, auditable, and efficient payment and reimbursement model. We have conducted case studies for each application to empirically prove its feasibility and test its scalability, stability, and efficiency.

6.2 Methods

The overall architecture, as shown in Figure 20, covers study finance and four different stages (e.g., study planning, study startup, study conduct, and study closeout) throughout the entire clinical trial process. This architecture is generalizable to all different clinical trials so that the participating site can use the same CTMS to manage simultaneous

clinical trials by switching trial IDs obtained by the sponsors while the registration on the blockchain-based CTMS remains constant. It is noteworthy to mention that CTMS may require more functions such as protocol development which are not included in our system design as the current procedures for protocol development are sophisticated enough [102] with no need to adopt a new approach such as blockchain to reinstate the existing process although most present tools can be integrated with our proposed blockchain-based CTMS without extensive arrangement.

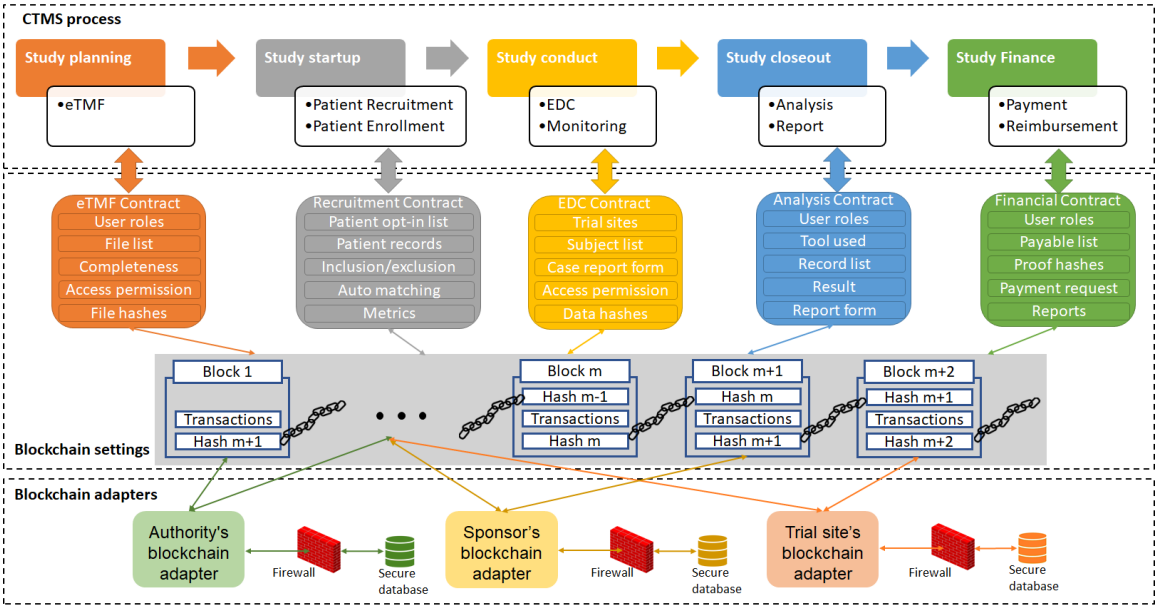

Figure 20. The overall architecture of five different clinical trial processes. Different applications are implemented by smart contracts defined from the blockchain initiation. Participating sites need blockchain adapters to interact with the blockchain system and the secure database protected by local health IT.

6.2.1 Environment Setup

Each blockchain adapter, as shown in Figure 21, installed the Ubuntu operating system, which in turn runs GoQuorum, an Ethereum-based Quorum blockchain client. The blockchain adapter will be added to the blockchain by the authority, and will be able to communicate with other blockchain adapters and the local secured database protected by

the health IT once the participating site obtains permission to join the system. Tools can be installed on the blockchain adapter and integrated with the blockchain through a Remote Procedure Call server. For example, a team of professionals such as medical experts, statisticians, clinical research coordinators, and medical writers can use blockchain adapters for protocol development. The existing tools can still be used as anticipated. The sole exception (limited to development scenarios) is the ability to store a log file in the blockchain after each use. In all other aspects, the users can take advantage of blockchain's unique features such as immutability to ensure file consistency, traceability to acknowledge the users who edited the file, and decentralization to improve the efficiency of working distributively without changing the existing legacy process. Each adapter has installed an InterPlanetary File System (IPFS) which is an innovative, peer-to-peer, distributed file system. Each file stored in IPFS is given a unique cryptographic hash for indexing and ensuring consistency. Compared to other distributed file systems, IPFS has shown great improvement in efficiency, scalability, and stability [103]. However, the design concept of IPFS lacks the capability of access control and tracking of file use [104] but this makes it a perfect match for

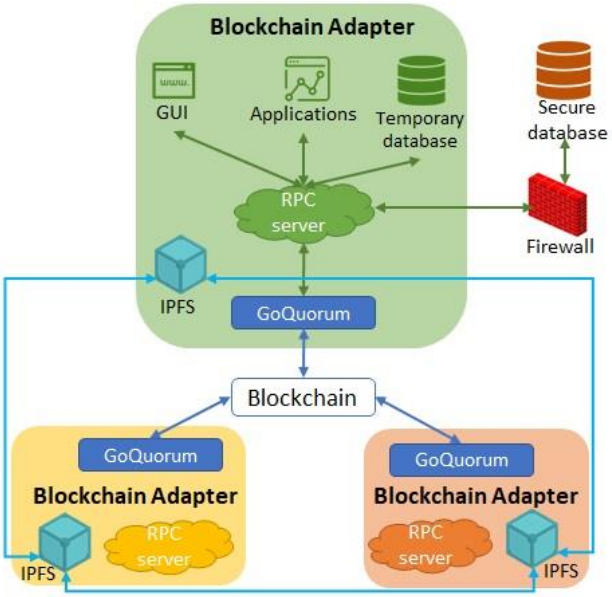

Figure 4. Blockchain adapters' design and connections. All adapters have the same setup with an RPC server connecting local applications and databases, an IPFS that connects to other IPFS on each adapter, and a GoQuorum API that connects to the blockchain.

blockchain. IPFS can be used as data storage while blockchain serves as a content management system.

A unique public-private key pair will be generated for each user such as subjects, investigator, sponsor, and others after the user registers a blockchain account through a site's blockchain adapter. Patients and potential subjects need to register on-site so that the administrators from trial sites can prove their identities and map their local patient ID to the blockchain account with their consent. A hash value of the public key, also known as the blockchain account address, will be used to represent the user's identity and the private key will be used as a digital signature. All transactions need to be signed by the sender's private key before they can be recorded into the blockchain. Each group such as the financial management team has an umbrella account in addition to separate individual user accounts, each of which that maps to the umbrella account for each member so that the whole group can share the permission once the authentication to the group is made. Potential subjects need to go to the trial sites to opt-in to the system and generate their blockchain account so that the trial site can verify their identities. Instead of memorizing the key pair, a username and password or biometric authentication mechanism can be used on a GUI for users to log in to the blockchain system.

6.2.2 Study planning

With the increasing adoption of electronic documents for clinical trials, planning, sharing, and managing the documents becomes increasingly critical and intricate [105]. The eTMF is a form of the content management system used to manage and collaborate in a timely fashion on essential clinical documents throughout the lifecycle of clinical trials. However, several persistent challenges such as inability to audit the unlocatable files,

inaccurate metrics for timeliness, quality, or completeness, inconsistency caused by loss or alteration of the information, and collaboration issues caused by different Trial Master File (TMF) standards exist in most eTMF designs [106]. Our eTMF design contains a smart contract used to control file access, validate file consistency, and manage collaboration on TMF development and the IPFS network used for file storage and file indexing. Figure 22 shows part of the source code of the smart contract in each function.

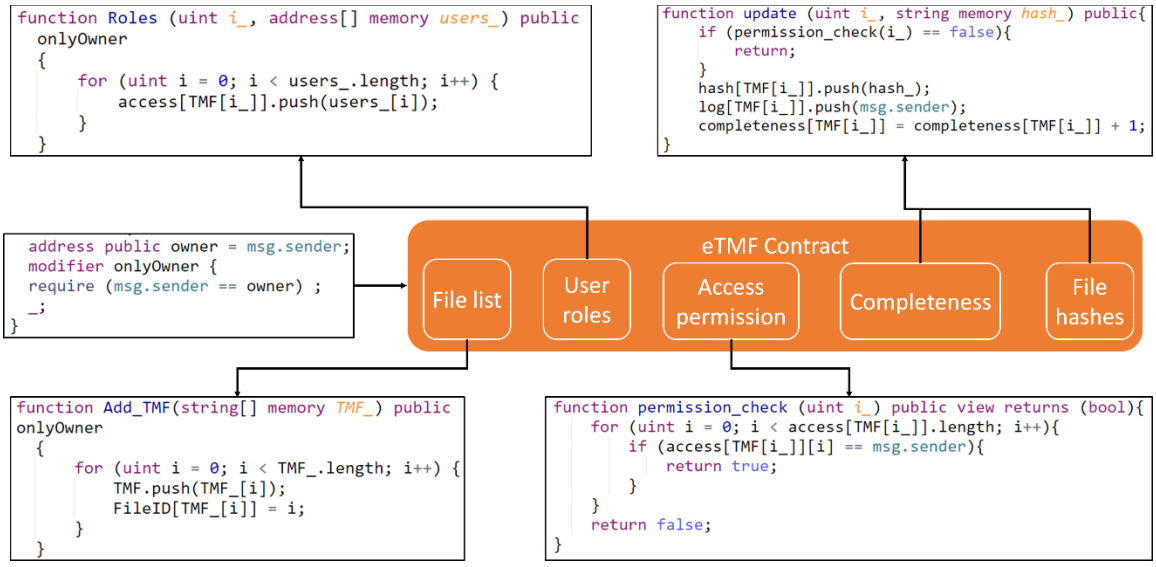

Figure 22. Part of the source code of the eTMF contract design. These codes show the main logic of each function. All smart contract functions are predefined and users can use GUIs to call the functions.

The TMF document list and other expected artifacts list need to be identified in the eTMF smart contract at the beginning of the study planning phase. Sponsors need to assign the files to team members so that they can work jointly by adding their blockchain accounts to the smart contract associated with the file ID from each TMF. All the TMFs are encrypted using OpenSSL and a randomly generated key pair before being stored in the IPFS [107]. All users can download the file from the IPFS using the file hash but only the

users who have permission from the sponsor can retrieve the decrypt key from the smart contract to decrypt the file. When a team member is working on a certain file, the blockchain adapter from the member's site will automatically send a flag to the smart contract to block other team members from working on the same file. When the team member finishes editing the file, the blockchain adapter will encrypt the new version of the file with a random new pair of keys, upload the encrypted file to the IPFS, obtain a new hash value from the IPFS, and send the decrypt key, the hash value, and negative flag to the blockchain to update the file registration information. The completeness metric (the percentage of completed artifacts of the expected artifacts) will be updated automatically.

Using blockchain technology for eTMF can provide the following unique features: (1) consistency - each version of a file will have a hash value stored in the blockchain, and any changes to the file will result in a mismatch of its new hash with the original hash; (2) traceability and auditability - each team member must work on the file sequentially so that any changes can be traceable to the editing user through blockchain transaction history (users can audit who has changed the file by checking the log files in the blockchain but only the sponsors, or the authority, knows the real identity of the user); (3) efficiency - using IPFS as file storage is efficient compared to other file transferring processes because team members can collaboratively work on the same file; and, (4) security - with blockchain's security setting, all transactions are considered secure so that only the recipients can receive the correct decrypt key for the file.

6.2.3 Study startup

After the study team has selected the trial sites and defined target enrollment metrics, clinical trials need to meet the recruitment goal. Patient recruitment has been recognized as a key to success. However, 86% of clinical trials fail to meet their recruitment goals on time. We have refined our earlier work which is a blockchain-based recruitment model using a smart contract for automated matching described in Chapter Five [84] for use under the CTMS study startup scheme, as shown in Figure 23.

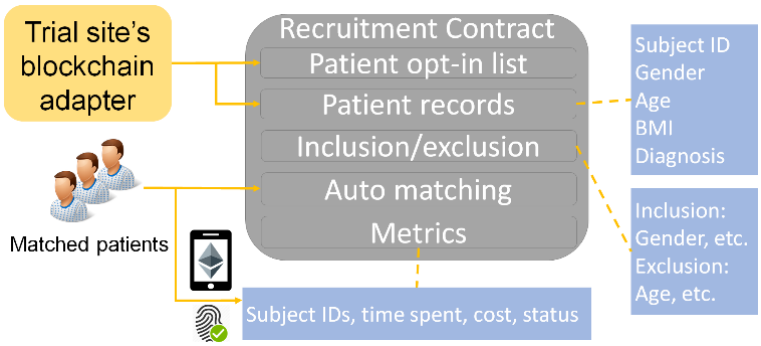

Figure 5. Trial sites must register subjects and input primary medical history to the smart contract. The smart contract will automatically send notifications to the matched patients asking for authentication through their mobile device using their fingerprint.

Users who want to participate in clinical trials must follow the same procedure outlined for patients and subjects; they also need to give permission for the use of their EHR for future matching purposes. The hospital's administrator needs to input the user's basic information into the Recruitment contract, including demographic information and primary diagnoses from past visits. Once the sponsor inputs the recruitment inclusion and exclusion criteria into the smart contract, the smart contract can automatically screen the potential subjects by matching the basic information. After that initial screening is accomplished, hospitals can perform precise matching by checking the matched users' full EHR. Once a user is fully matched, the sponsor will send a transaction to the user to ask

for enrollment. Future on-site visits are still needed, but the enrollment process can be operated by sending out the consent form and asking the user to sign using their private key, which will send a confirmation transaction to the smart contract. The smart contract also contains personalized metrics such as time consumption, cost, and retention used for evaluating the performance of the team performing the recruitment process and the timeliness of decisions to increase productivity.

The features of blockchain technology are a great fit for the needs of recruitment and enrollment for the following reasons: (1) transparency can improve the awareness of the clinical trials for patients, (2) auditability ensures the legitimacy of clinical trials, (3) anonymity protects patient's privacy, (4) asymmetric encryption eases the process for patient enrollment, and (5) the automated matching mechanism operating via a smart contract can significantly reduce the time-consumption of recruitment.

6.2.4 Study conduct

During the conduct of clinical trials, data collection is one of the most important processes for the evaluation and monitoring of aspects of the experimental condition (e.g., a drug effect). Compared to the traditional paper-based Case Report Form (CRF), which served only the purpose of recording information, EDC systems are used to collect data electronically, to reduce data errors, to improve the efficiency of the collation process, and to enable faster access to the data. However, there are several challenges faced by both the paper form and an EDC system such as security concerns, data inconsistency, and untimely (slow) data input. All clinical trials are monitored, which is a process of data monitoring and safety monitoring. The Data and Safety Monitoring Board (DSMB) is a group of professionals from different fields such as biostatistics, medicine, and ethics, who monitor

patient safety and treatment efficacy. The legacy data monitoring method is Source Data Verification (SDV) which is resource-intensive and can take up to 30% of the total clinical trial budget. We have designed the EDC contract to effectively collect data, reduce the need for SDV, and monitor patient safety persistently.

After subjects send a consent to the blockchain from the recruitment phase, the system administrator from each trial site must register them in the subject list in the EDC contract to map their blockchain account to the trial ID and their local patient IDs. Sponsors need to customize their CRF on the smart contract to identify the data fields used for collection. After each subject's site visit, the investigator needs to input the records into the eCRF. The records will then be automatically encrypted, hashed, and stored in IPFS by the site's blockchain adapter. Figure 24 shows the GUI for investigators and the hashes returned by the IPFS after storing the encrypted data and Figure 25 shows the encryption, storing, and retrieving process after the data is input through the GUI. The smart contract will validate whether the trial site has permission to store the subject's data after which the visit ID and decrypt key will be sent through Quorum blockchain's private transaction. This ensures that the data contained in the private transaction are encrypted and only the recipient can decrypt using their private key, or the information can be made available to the sponsor by the site's administrator. The sponsor's blockchain adapter will automatically retrieve the decrypt key and hash from the blockchain, decrypt the records and hash the records to compare with the hash stored in the blockchain. Mismatching hashes will create an alert to the trial site and for the sponsor so that further investigation begins. This can eliminate the data inconsistency caused by falsification. However, most EDC needs manual input and human data errors can also cause a data inconsistency issue.

We have implemented a data extraction application on each blockchain adapter to automatically extract CRF required data from the visit records before storing it in the secured EHR database to reduce the risk of human error. However, most CRFs have partial data fields that are trial oriented and not included in the EHR, meaning that manual input is still needed. Although blockchain's immutability features were intended to be designed as unchangeable for all records, some modifications may still occur due to unintentional human error. However, the updated (erroneous) records cannot replace the previous input and will contain a pointer to the former hash of the data record for future validation. In this blockchain-based CTMS system, safety monitoring relies on the investigators to report through EDC so that the safety monitor team may evaluate only true issues of data and safety.

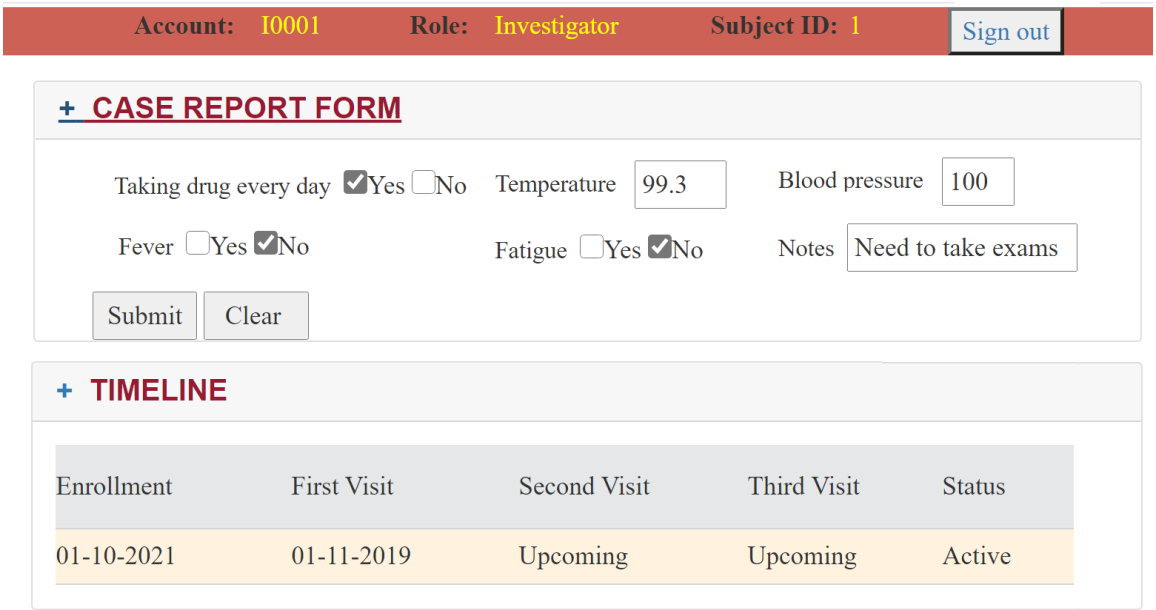

Figure 24. The GUI for sponsors contains a sample eCRF coded through the smart contract and a sample timeline for the subject. After submitting the input, the data will be retrieved by the blockchain adapter.

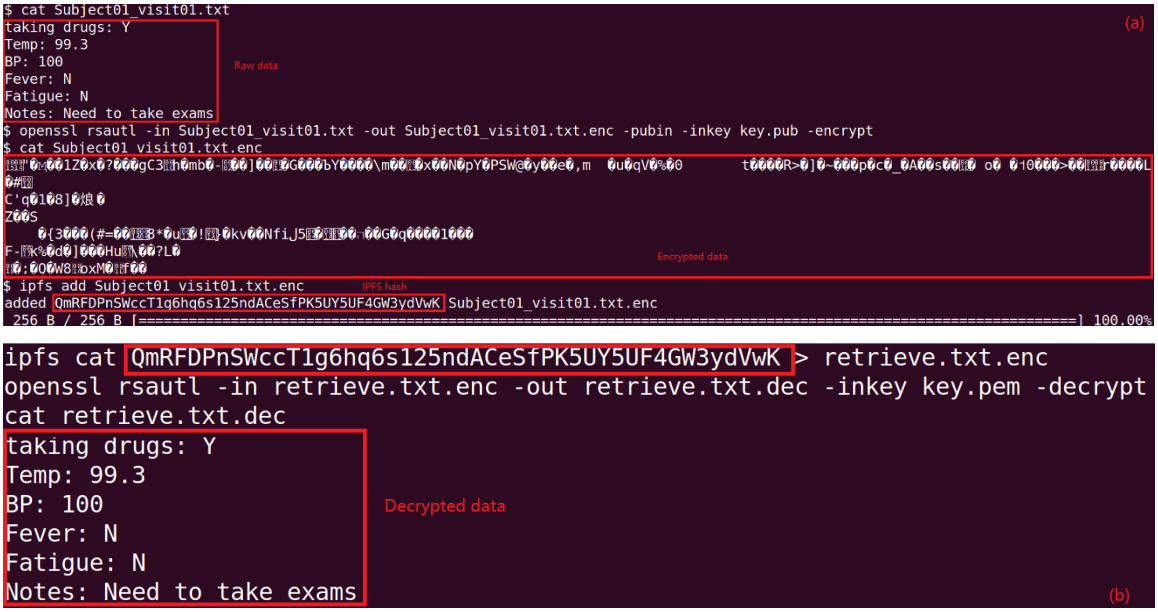

Figure 25. (a) The investigator's blockchain adapter retrieves the data through GUI, encrypts the data using the investigator's public key, and stores the encrypted data into IPFS. (b) The sponsor's blockchain adapter retrieves the encrypted data through IPFS and decrypts the data using the private key.

In this module, using blockchain and IPFS for EDC has the following benefits: (1) immutability ensures data consistency from the data input through to the data analysis to reduce the need for SDV, (2) traceability improves the auditability as to who, when, and how the records were changed, (3) the efficiency of IPFS permits fast data retrieval, and (4) the security property of blockchain protects patient privacy and data security. With the automated extraction mechanism added to the blockchain adapters, the efficiency and accuracy of the data collection process have been significantly enhanced.

6.2.5 Study closeout

When the last subject completes their site visit, the clinical trial will enter the closeout phase. There will be a closeout checklist that can be collaboratively completed by sponsors and the team using the eTMF. The clinical trial database can be locked to prevent future changes after the final data validation. Statistical analysis must be conducted to evaluate the outcome of the clinical trial. In the blockchain-based CTMS system, we have

created several R scripts for several statistical models in each blockchain adapter and added the names of available statistical methods in the smart contract. The statisticians can use the existing script or use their preferred statistical tool to analyze the final data, after which they can generate the final statistical report. The source code must be encrypted and stored in the IPFS for validation purposes. The team members or the authority can request the decrypt key from the sponsor and reproduce the results using the source code and clinical trial data.

Barriers to analyzing clinical trials are mainly those of selective reporting [108], incomplete reporting data [109], and a lack of appropriate statistical methods [110]. Blockchain provides solutions to the challenges in this stage through its immutability and auditability features which help to ensure the reports' completeness. The analyzed data and applied methods will store a log file in the blockchain so that the study group and the authority can reproduce the result at any time to validate the completeness and audit appropriateness of the analyzing methods.

6.2.6 Study finance

There are numerous components that can add to the cost of a clinical trial such as regulatory services, start-up, and medical writing, all of which can make financial management challenging. In this module, we use payment and reimbursement to the trial sites and patients [111] as an example of the potential use of blockchain technology as a financial management tool. The validation of the payment or reimbursement requests as to when and how the recipient gets paid is a time-consuming process so that in-time payment is challenging [112]. In this module, we have designed a smart contract and a collaborative validation network in the blockchain-based CTMS.

Before the clinical trial begins, the study team should define a list of payable entities (as well as payable items) and input this list into the smart contract. This can standardize the payable items and reduce the risk of hidden fees. Each trial site may have different rates for the same payable item. The rates must also be defined through a smart contract only accessible by the sponsor and the trial site. Compensation to the patient is normally based on the subject's time required to participate. After each visit, the trial site needs to send a request transaction containing the time spent and the payable items to the blockchain, store the encrypted proof in the IPFS, and send the decrypt key and hash to the sponsor. The clinical trial financial management team can validate the proof and send the payment request to the sponsor. A transaction that contains a payment receipt will be sent from the sponsor to the trial site and mark the status of the request as paid in the trial site's GUI. The payment to or reimbursement of the trial site has a similar process as trial sites send request transactions that contain payable items to the sponsor and wait for the approval. However, the payable items may not cover all requested payments. Trial sites need to follow the same request process with "additional items" in the payable items. Sponsors can collaborate to validate the proof and price the additional items to make the payment.

Using blockchain technology for financial management has the following benefits: (1) a customizable charging standard for different trial sites as long as the sponsor agrees (all payable items and rates are preferred to be defined in the smart contract for an expedited validation process); (2) the traceability feature ensures all requests and payments are traceable by the requester and the recipient (all the proof needs to be stored in the IPFS); (3) the immutability feature ensures the request, the payment, and the proof of payment are

not modifiable after the payment is made; and, (4) the security property of blockchain protects users' privacy.

6.3 Case study

We have implemented the blockchain-based CTMS and installed it on six blockchain nodes representing one authority, two sponsors, and three trial sites. Each blockchain node has been converted into a blockchain adapter. We have generated two clinical trials with 1,000 subjects on each trial site for each study. This case study simulates the data collection process described in the Study Conduct section.

We have designed a sample eCRF through the smart contract. A script is created to mimic the data capture process: (1) to randomly generate data for the data fields defined by the eCRF from the three trial site adapters; (2)the trial site adapters encrypt the data file using a random public key, store the encrypted data file in the IPFS, then obtain the hash value and send decrypt key and hash value to the sponsor through a private transaction; and, (3) the sponsors' adapters retrieve data from the IPFS and decrypt the data files. We have run the script on each subject from each blockchain adapter every second for an hour. There are 1.2 million transactions written into the blockchain with an average latency of 1.73 seconds and transaction per second (TPS), a key measurement of blockchain scalability, of 335.4. The remainder of the transactions were held in the buffer to push into the blockchain sequentially. It took nearly 18 hours to send 21.6 million transactions generated by the script into the blockchain with a 100% success rate. Figure 26 shows blockchain performance after submitting 2000 transactions simultaneously. The average TPS is 458.9 but decreases to stable gradually during the simulation. TPS is not associated with the block generation time from our simulation results.

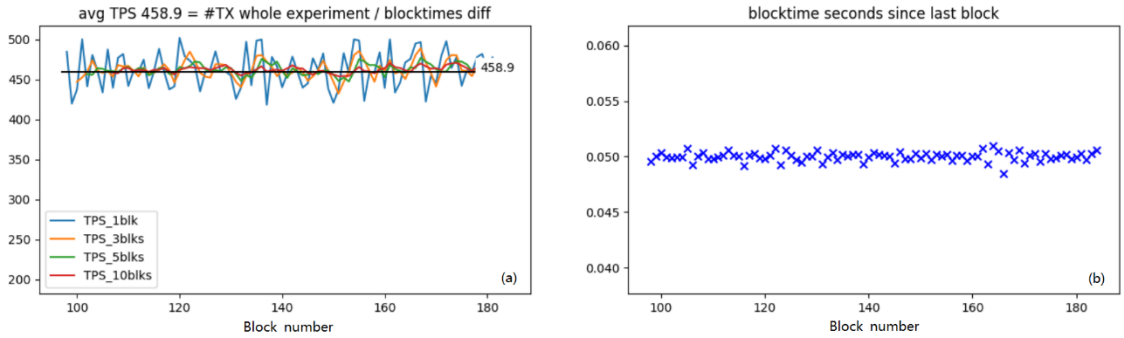

Figure 26. Scalability and stability test result of first 2000 simultaneous transactions. (a) TPS values calculated using every 1, 3, 5, and 10 blocks. (b) time consumption of generating a new block.

Since script #3 is purely off-chain, the stability is based on the performance of the IPFS and the specifications of the adapter's devices. We have not included script #3 in our stability test since many researchers have proved the performance of the IPFS [113]. To test system robustness, we have manually shut down the sender's blockchain adapter after the transaction and found that the recipient can still retrieve the data.

6.4 Discussions

In this chapter, we described a blockchain-based CTMS that covers four different stages of clinical trials. Through our simulation process, our system empirically proved the feasibility of the architecture. Comparing to the scalability test on the Ethereum blockchain from our previous research, Quorum blockchain shows an overall better performance. The unique contribution of this work is exploring the benefits of blockchain technology in targeting the needs of CTMS. This covers several essential functions (each of which is a part of the clinical trial process) using a distinctive blockchain adapter design to support an efficient, secure, traceable, transparent, and auditable management system.

CHAPTER SEVEN – VIRTUAL CLINICAL TRIALS

We have published this chapter in the Annual Symposium proceedings in 2020 [114]. This chapter adopts its main contents with minor modifications.

Zhuang Y, Sheets L, Gao X, Shen Y, Shae Z, Tsai JJP, Shyu CR. Development of A Blockchain Framework for Virtual Clinical Trials. AMIA Annu Symp Proc. 2020

7.1 Background

Traditional clinical trials face multiple challenges related to patient recruitment [115], patient engagement [116], and cost [117], as listed in Table 7. Virtual clinical trials (VCT) represent a relatively new approach to conducting clinical trials solely through digital health platforms to make participation transparent for subjects [118]. Compared to traditional clinical trials, VCTs have three major advantages: 1) improving recruitment by maximizing accessibility to opportunities for patients to participate at their homes rather than traveling to the clinical trial sites, which is particularly important for patients with mobility issues or who live far from trial sites [119], 2) keeping subjects engaged throughout the study and providing a patient-centered approach through real-time data collection [120], and 3) preserving cost-effectiveness by minimizing money and time patients spend traveling to clinical trial sites [121] and reducing the cost of managing clinical trial sites for sponsors although assistance in the registration and education of patients by clinical sites is still required [122]. Pfizer pioneered the VCT using web-based platforms to conduct randomized clinical trials [118]. The VCT has had limited success and has failed to recruit enough subjects [123], but has shown a degree of success on the feasibility of home-based VCTs by distributing drugs in a double-blind study and using

digital health platforms for data capture. Despite numerous benefits, VCTs remain the exception rather than general practice due to several persistent challenges: 1) an immature recruitment model reaching insufficient numbers of patients [121], 2) patient privacy and data security concerns caused by sharing health information over the Internet [119], 3) technical challenges in developing a distributed patient engagement and monitoring platform [121], and 4) cultural barriers such as skepticism about the technology or lack of computer literacy [119]. Considering these challenges and features, blockchain technology could be a perfect match for VCT.

Table 7. Traditional clinical trials challenges and causes.

Challenges	Patient recruitment	Patient engagement	Cost
Causes	1. Lack of awareness 2. Distrust of the trials 3. Protocol limitation 4.Inaccessible clinical sites	1. Inaccessible clinical sites 2. Lack of literacy on the clinical trial 3. Passive role in clinical trials	1. Staff and administrators 2. Site monitoring 3. Site retention

Applying the original blockchain for the VCTs can protect data security and patient privacy, ensure data consistency, and publicize the information to all the users. The Ethereum blockchain keeps all original blockchain features and adds a new function called a “smart contract” which makes blockchain more suitable for healthcare applications. A smart contract is a self-executing protocol running on blockchain to regulate transactions[21]. Smart contracts in Ethereum can be coded to solve any computational problems since they are coded using Solidity, a Turing-complete language[124]. For example, a data analytics tool can be coded inside the smart contract to do real-time

anomaly detection during the VCT. Sponsors or the VCT authority (such as the FDA) can ensure timely medical assistance for patients.

This work demonstrates the feasibility and robustness of applying blockchain to VCT by implementing a functional blockchain system with multiple smart contracts. We have also conducted simulations of the VCT process, from patient recruitment to persistent monitoring of anomalous patient outcomes.

7.2 System Design

We implemented a private Ethereum blockchain framework, shown in Figure 27, consisting of three different modules to simulate the VCT process: (1) *Patient Recruitment* module: a smart contract that automatically matches potential subjects, asks matched patients for consent to join the VCT, and generates a specific trial contract for each VCT that is only accessible to the participants. (2) *Patient Engagement* module: a smart contract to allow patients to input data and interact with clinical trial sponsors or principal investigators, and (3) *Persistent Monitoring* module: a smart contract to persistently monitor anomalies through the analytics tool, either installed on the sponsor's node or embedded inside the monitor contract.

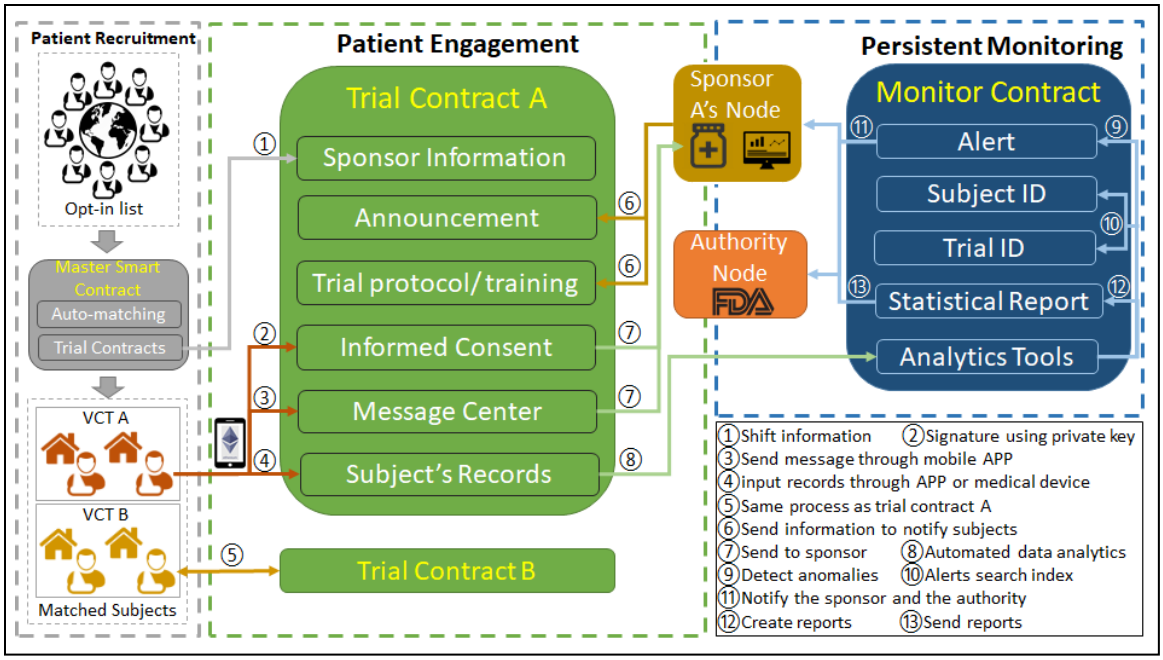

Figure 27. Blockchain framework with multiple smart contracts across three different modules: patient recruitment (based on prior work), patient engagement, and persistent monitoring.

7.2.1 Environment Setup

Our blockchain system requires the authority, each VCT sponsor, and each clinical site to provide a blockchain node. We have packaged the setup process into an executable program for the authority node to initiate a unique private blockchain. All other parties receive permission through a JSON file that contains the information of the private blockchain from the authority, then deploys the file using the user's blockchain node. After the setup procedures, all the nodes are in the same blockchain network and can communicate with each other. When a blockchain node joins the blockchain, a blockchain account with a public/private key pair is automatically generated for the system administrator. Patients go to clinical sites to opt in the system to prove their identities, and to get orientation on operating the system. The clinical site administrator creates a blockchain account for the patient and links their information to the blockchain account

with their consent for future subject matching. Patients can add biometric information to their accounts for future authorizations.

Two smart contracts, a “master smart contract” which is mainly used for patient matching, patient recruitment, and trial contract generation, and a “monitor contract” which is used for persistent anomaly monitoring during the VCT, are automatically deployed after the authority builds the blockchain. Web-based GUIs are implemented on each blockchain node for users to interact with the smart contracts. Patients can use GUIs through their smartphones or home computers to participate in VCTs.

7.2.2 Patient Recruitment

When patients opt-in to the system, their primary visiting histories are input to the smart contract by the clinical sites, including demographic information and primary diagnosis and treatments from each visit. These records are used for primary subject matching. The authority node automatically extracts the inclusion/exclusion criteria from the clinical trial database and inputs them to the master smart contract after a VCT is approved. Patient matching is a two-step process: first, the master smart contract automatically matches potential subjects using their primary visiting histories; second, clinical sites help the sponsor find precise matches with consenting patients. The smart contract requires patients to provide consent to share their full records with the sponsor if they are primarily matched. If the patient is fully matched the smart contract sends a notification to the patient asking for consent to join the VCT and permitting them to access the trial contract. This recruitment module can save time for patients by only checking the potential matched clinical trials rather than browsing all recruiting trials. The VCT’s information is input by the authority which eliminates the concern of scam trials and

"spearphishing" for the patients. As shown in Figure 27, the potential subjects for each VCT can be matched from registered users all over the world through the master smart contract. Matched patients can engage in the VCT through the trial contract automatically generated by the master smart contract. More details of implementation and simulation results can be found in Chapter Four [125].

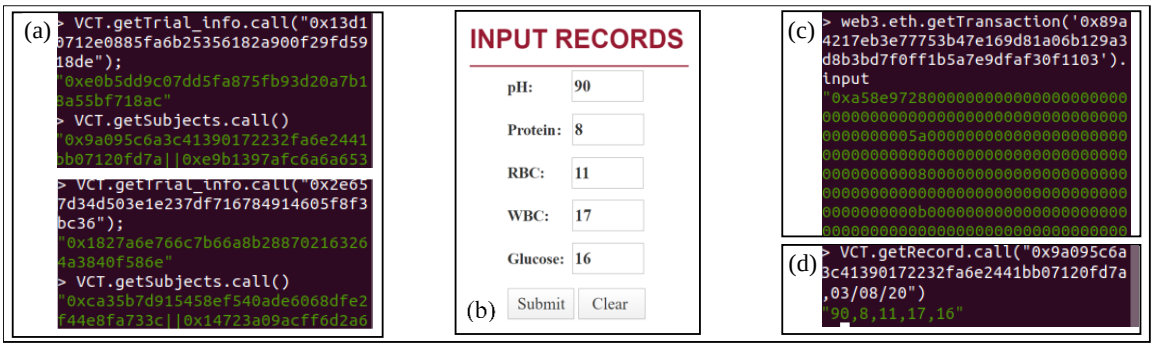

Figure 28. (a) Trial contract information retrieved by the trial sponsor through the blockchain console, (b) GUI for patients to input data manually or from a connected medical device, (c) scrambled patient records retrieved by unauthorized users through the blockchain console, (d) the patient record retrieved by the sponsor through the trial contract function by inputting the patient's blockchain ID and the input date.

7.2.3 Patient Engagement

The master smart contract automatically generates a random ID for the VCT and a pre-coded trial contract if the VCT is approved. The trial contract's information is stored in the master contract for management. Only the sponsor, the authority, and matched subjects are granted access to the trial contract. Other users will not see the contents of the trial contract. We supported two simultaneous ongoing VCTs in our simulation. Figure 28(a) shows the information included in two trial contracts and their access list within our simulation. The upper figure shows the trial contract address and subject list for VCT 89938. The lower figure shows the information for VCT 71115. Only the listed users have access to the trial contract. Other users are not able to execute any functions through the

trial contract. These figures show the result of inputting commands through the blockchain console using the corresponding sponsor's blockchain account. The sponsor inputs the informed consent form to the smart contract for the subjects to sign digitally using their private keys. After signing the consent form, patients can have access to the other functions of the trial contract, such as getting the trial protocol and training, getting announcements from the sponsor, and sending messages to the sponsor or principal investigators, and inputting health information, as shown in Figure 27 ②-⑧. Patients input their measurement records to the blockchain through the web-based interface, mobile app, or even automatically from a connected medical device. Figure 28(b) shows the GUI for patients to input their measurement records. After the record is input through the GUI, it is automatically sent through a transaction in the blockchain. Patient's records are automatically encrypted by the sponsor's public key and can only be decrypted by the sponsor's private key. The sponsor can automatically view this decrypted record through the GUI for trial management. However, since all the transactions are publicly auditable, other users can still see the transaction through the blockchain console. Figure 28(c) shows the patient record as viewed by an unauthorized user; without the sponsor's private key the input data is unreadable. Figure 28(d) shows the patient records retrieved by the sponsor from the blockchain backend console by inputting the patient's blockchain account and the input date. This process only shows the security setting for the data exchange platform. Users can always choose the user-friendly interface to operate the system.

Blockchain in the patient engagement module provides the function of an electronic data capture system that is designed for the collection of clinical data during clinical trials [126]. Patients can input their data through the GUI in the format defined by the sponsor.

All records are securely stored in the blockchain and can only be retrieved by sponsors. This blockchain framework can be a uniform platform used for all VCTs through the unique trial contract without third-party management. Sponsors can develop their VCT applications to improve patient engagement. There are many checklists to evaluate the quality of the application design, such as the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) [127]. Blockchain does not require the applications to following certain standards. However, blockchain enhances patient engagement in the following four aspects, 1) informing patients [128]: any VCT related tools deployed on the blockchain through trial contracts are accessible for participants. Patients stay tuned for timely documentation and education through blockchain, 2) engaging patients[129]: patients can access their trial records anytime through the trial contract. The trial contract acts as a patient portal to provide the functions for patients to interact with principal investigators and sponsors, 3) empowering patients [130]: in this platform, patients are not only data points but also become involved with the trial by interacting with the sponsor to express their feelings about the trial through anonymous secure messages using the trial contract. Patients become an active role instead of the passive recipient after sponsors received and took into consideration of the patients' feedback through blockchain, and 4) partnering with the patients [131]: with the abovementioned engagement, patients will obtain experiential knowledge about their personal health condition. Sponsors can include patients as advisory board members [132], steering committee members [133], or even co-investigators [134] to improve the future study design and achieve better clinical trial outcomes.

7.2.4 Persistent Monitoring

This module contains two parts to persistently monitor the subjects' physical conditions and the VCT status. The first part is alert creation when anomalies are detected and the second part is statistical report generation after the clinical trial is finished, as shown in Figure 27 ⑨-⑬. Subject records can be retrieved by the sponsor with a log file stored in the blockchain. The monitor smart contract will automatically use the analytics function to analyze the patients' records inside the blockchain when a new record is input to the system. The monitor smart contract generates a periodic report of all patients which is viewable by the authority. As all the transactions are publicly auditable, the analytics process is also under the authority's surveillance, which eradicates the concern of tampering with data in the sponsor's final report.

When an anomaly occurs in a patient's record, such as lab test results outside pre-defined threshold values, as shown in Figure 29(a), or an abrupt change of previous records, or a severe vital sign measurement values, the smart contract automatically sends an alert to the sponsor for immediate action. This action may include sending a notification through blockchain to the patient to ask whether the change is caused by human error, having a clinician communicate with the patient to provide medical help, or even calling an emergency telephone number after locating the patient through the registered clinical site. Figure 29(b) shows the alert received by the sponsors and Figure 29(c) shows the patient's GUI when they received the alert. This protects patients' safety while obtaining their acknowledgment of their physical condition during the VCT.

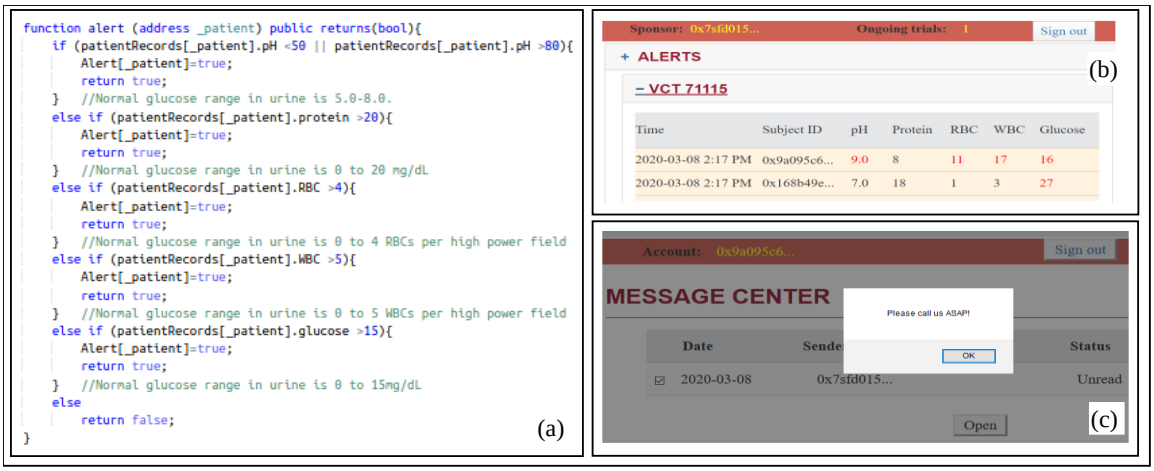

Figure 29. (a) an alert system defined in the smart contract to detect abnormal values, (b) the GUI for sponsors to receive alerts during the clinical trial. Abnormal values are marked as red font automatically, (3) the GUI for patients to receive messages from sponsors.

The monitor contract can also trigger an analytics tool installed on the sponsor's node and keeps a log of data use in the blockchain. The tool can be a basic comparison model as shown in Figure 29(a), a simple statistical model to summarize the effectiveness of the new treatment, or even a sophisticated machine-learning model with AI components to detect comorbidities and predict outcomes depending on the needs of the sponsor.

7.3 Simulation and case study

We have simulated the patient engagement process using our system to test the feasibility, security, and efficiency of this system. We deployed six physical nodes using Intel NUC machines representing one authority node to initiate the private blockchain environment, two sponsor nodes to start two different VCTs simultaneously, and three different clinical site nodes for patient registration. We set up 1,000 synthesized patient accounts on each clinical site's node and randomly chose 3,000 patient records from the MIMIC-III (Medical Information Mart for Intensive Care) Clinical Database for our simulated patients[135]. We used MIMIC-III for this simulation because it contains lab test data for each patient visit. The commonly encountered lab values selected for this

simulation were urine tests for pH, protein, red blood cells, white blood cells, and glucose. For the purpose of this simulation, we assumed that the subjects were performing the tests at home with a medical device that automatically inputs the data to the blockchain. Each simulated VCT randomly enrolled 1,500 patients distributed over three clinical sites. We created a script to send records from each patient's account to the smart contract for the corresponding trial contract. The script recurrently sends 10 patient test results to the smart contract every second for 24 hours, which simulates 288 days of VCT with 864,000 transactions in total. The average time for the sponsor to receive the data is 8.31 seconds.

We have created pre-defined thresholds for each test item in the smart contract as shown in Figure 3(a). The case study tests the alert system provided by the persistent monitoring module; after each record is input by the patient, the monitor contract automatically detects abnormal values and sent an alert to the simulated sponsor. There were 95,934 alerts in total, at an average of 1.17 seconds after the data was sent to the smart contract. By the end of the simulation, the monitor contract had generated statistical reports for two VCTs for the sponsor and the authority to review. Statistical reports can also be generated for individual subjects to have a better understanding and engagement of the VCT. Subjects can check the statistical reports based on different test items throughout the VCT so that they can acknowledge the long-term trend of their physical condition, as shown in Figure 30(a). Comparing to the overall statistical report, as shown in Figure 30(b), subjects can have a better understanding of the VCT outcomes such as efficacy and safety of the treatment. These graphs were generated by passing the values from the blockchain to the Google chart application programming interface from the monitor contract.

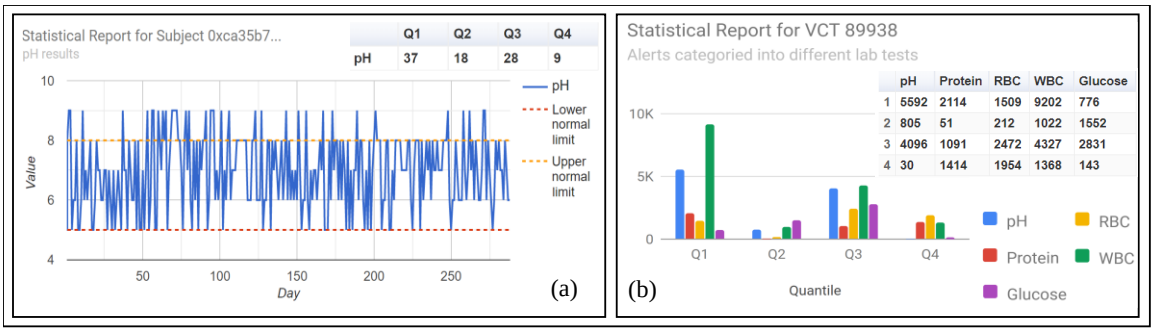

Figure 30. (a) Statistical report of pH test of an individual subject with normal limit labeled. Subjects can acknowledge the long-term trend based on occurring anomalies and the comparison of the VCT report, (b) the statistical reports on the total alerts received on abnormal values in VCT 89938. The alerts have been divided into four quantiles of the studying period to show change over time.

7.4 Discussions

Our simulation shows the feasibility, stability, security, and robustness of applying blockchain technology to achieve better patient engagement and persistent monitoring during VCTs. Compared to traditional clinical trials, our simulation proves blockchain-based VCTs have a more efficient and secure patient recruitment process, a more patient-centered engagement platform that uses patient involvement to move the clinical trial forward by distributively collect health information, a better-automated data analytics tool embedded in the system that can detect anomalies in patients' records in real-time, and better security to minimize the risk of tampering with records to manipulate statistical reports at the end of the clinical trial for subjects, sponsors and the authority to have a better understanding of the VCT outcomes. The alerting system shown in the simulation empirically proves the practicability of real-time detection of anomalies. In the real VCT cases, subjects with different medical histories may have different reactions to the treatment so that subjects under the normal physical condition may have some test values exceeding the normal limits. To be specific, the normal limit of the glucose value in urine is 0-0.8mmol/L. However, there is a high possibility for a newly diagnosed type II diabetes

patient to have glycosuria in the short term under the current treatment as taking metformin or insulins. Therefore, the sponsors or PIs can personalize the alerting mechanism by adjusting the trigger limits through the smart contract depend on the treatment effect on different populations. Since blockchain is a distributed ledger technology, subjects can participate from any place at their convenience, removing the burden of traveling to clinical sites.

Multiple smart contract settings optimize the clinical trial process and management. The master smart contract registers all patients who opt-in to the system, generates trial contracts for VCT management, and automates patient matching and recruitment. Patients are added to different trial contracts to conduct future VCTs after being matched. Trial contracts are accessible only to the participants in order to protect data privacy. Patients can use a patient-centered interface to engage with clinical trial sponsors through trial contracts. Rather than only being a data point in the system, patients contribute their conditions and insights to sponsors and the trial authority. The monitor contract persistently monitors the physical condition of each patient and the effects of the treatment. Subjects receive timely medical assistance if they report anomalous measurements or a health emergency during the VCT. Using the monitor contract to general the final report eliminates the concern of tampering with trial results.

The simulation mainly simulates the data input process, which is the core process in the VCT. To test stability and efficiency, we synthesized multiple patient records with test values in each record. This meant that each transaction contained lab test results so the final report was not statistically interpretable. In real-world scenarios, the monitor contract can generate contingency tables to test treatment effectiveness by using the chi-square test.

Machine learning tools can also be installed on sponsor nodes and can be triggered by the monitor contract to provide more powerful real-time analysis, such as adverse-effects detection and prediction of possible outcomes.

The main limitation of this work is the patient recruiting process still relies on clinical sites to prove patients' identities, provide system literacy, and perform precise matching using patients' records. The system needs a reward mechanism to provide an incentive for clinical sites to participate. A business model must be developed with the cooperation of sponsors, clinical sites and the authority. The simulation empirically proved the feasibility of using the proposed blockchain framework and generic smart contract functions for VCT applications. The framework does not restrict the application development for different VCTs. Sponsors can develop unique VCT tools for the participants using the blockchain framework. Evaluation of the usability of VCT applications can be conducted through the trial contract.

CHAPTER EIGHT – CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

The objective of this dissertation is to utilize blockchain technology to improve the clinical trial process. To achieve this objective, we have designed and implemented multiple blockchain models based on the challenges of current clinical trial processes. Our system design, implementation, and simulation results demonstrate the potential of blockchain for clinical trial applications, and we suggest that this should serve as a notice for the health IT community to take this emerging technology into consideration. The unique contributions of this work provide feasible solutions for clinical trial challenges:

(1) Converting traditional blockchain nodes into blockchain adapters to abide by the local healthcare facilities policies, connect the on-chain and off-chain activities, and provide graphical user interfaces for users to have a better visualization of the interaction with the blockchain system.

(2) Augmented, generalized, layered architecture design, which fits a wide spectrum of healthcare applications, offers health technology community blockchain features for application development without requiring developers to have extensive experience with blockchain technology.

(3) Unique smart contract settings for different clinical trial applications such as an auto-matching mechanism for clinical trial recruitment, customizable EDC platform, and personalized data segmentation tool for patient-centric HIE.

(4) Large-scale simulations are conducted to evaluate the feasibility, stability, and robustness of the proposed blockchain models for clinical trial applications.

In Chapter Two, we introduced blockchain technology and its key features such as transparency, auditability, traceability, data provenance, consensus, peer-to-peer capability, distributed ledger, robustness, security, anonymity, and smart contracts. The concept of the blockchain adapter fulfills the scarcities of blockchain technology. Assumptions of the system design are also clarified in this section.

In Chapter Three, we have implemented an augmented, three-layered blockchain architecture that offers generic data coordination functions for healthcare application development without requiring developers to have extensive experience with blockchain technology. The three layers, from bottom to top, are as follows: (1) incorporation of fundamental blockchain settings and smart contract design for data collection; (2) interactions between the blockchain and health care application development environment using Node.js and web3.js; and (3) a flexible development platform that supports web technologies such as HTML, HTTPS, and various programming languages. An HIE example application was developed in our design to demonstrate the feasibility of the layered architecture.

In Chapter Four, we have proposed a feasible blockchain solution to patient-centric HIE barriers such as security and privacy concerns, data inconsistency, and timely access to the right records across multiple healthcare facilities. With the smart contract design, our blockchain model protects data security and patients' privacy, ensures data provenance, and provides patients full control of their health records. By personalizing data segmentation and an "allowed list" for clinicians to access their data, this design achieves

patient-centric HIE. We conducted a large-scale simulation of this patient-centric HIE process and quantitatively evaluated the model's performance.

In Chapter Five, a novel blockchain model is introduced containing multiple trial-based contracts for trial management and patient engagement and a master smart contract for automated subject matching, patient recruitment, and trial-based contracts management. Through the simulation process, the proposed blockchain model shows the capability of tackling clinical trial recruitment issues. Using the master smart contract to match patients and trial-based contracts to manage the clinical trials can optimize the recruitment process to allow fortime-saving, identifying all potential subjects, patient empowerment, and the authority's surveillance. Trial-based contracts can be used for EHR collection for subjects during the clinical trial. Blockchain features can ensure the data provenance is clinical sites, data consistency over time, data security that can only be decrypted by certain users, and patient privacy.

In Chapter Six, we have designed a blockchain framework for CTMS which covers the four stages, study planning, study startup, study conduct, and study closeout, throughout the clinical trial process. We have implemented an application for each module as collaborative eTMF, automated matching patient recruitment and simplified enrollment, precise EDC, and a reproducible data analytics platform. An efficient payment and reimbursement model for study finance management was also added to the blockchain-based CTMS. The innovative IPFS integration design strengthens blockchain features and increases the feasibility of utilizing this technology for clinical trial applications.

In Chapter Seven, we have implemented a blockchain architecture to achieve VCT, an innovative approach to conducting clinical trials. Compared to traditional clinical trials,

VCTs have shown multiple benefits in removing burdens from patient engagement. We have developed a comprehensive framework that covers the whole process of VCT as patient recruitment, patient engagement, and persistent monitoring modules, and tests its feasibility, stability, and security through simulation processes.

8.2 Limitations

The main limitation of this project is the setup required at each healthcare facility. Each healthcare facility is required to provide at least one node to the blockchain and complete the process of converting servers into blockchain adapters with local health IT compliance. Secondary limitations include the dependence of the model's performance on the blockchain nodes' properties, the potential need for patients to provide blockchain nodes for data generated by the Internet of Things (IoT) devices, and the need for facilities to agree on an interoperability standard such as Fast Healthcare Interoperability Resources (FHIR).

Another limitation is scalability constraints from the blockchain protocol [48]. Ethereum can handle roughly 13-15 transactions per second as of today. The Quorum blockchain has better scalability but the total number of transactions may exceed the limit at any moment. Our simulation provides a partial solution to this limitation by spacing the transactions as queuing them up by the adapters. This allows us to monitor the speed at which transactions are written to the blockchain and to buffer the backlog.This spacing solution is to ensure the successful delivery of the requests under the intrinsic scalability constraint existing in the current blockchain protocol.

8.3 Contributions to Informatics

This work aims to utilize the unique features of blockchain to solve the challenges of the current clinical trial process with the following significance:

1. *This study would provide a practical, generalizable blockchain framework for the healthcare community to adopt blockchain technology and develop other healthcare applications in a timely fashion.*

Comparing to the traditional blockchain architectures, we have proposed a practical generalized blockchain architecture that provides generic functions and methods for the application developers to securely collect data from different sources without requiring proficiency in blockchain technology. The architecture retains all the blockchain features and can be used as the foundation of our implemented and future applications.

2. *This study would utilize blockchain features and implement multiple applications that potentially solve current clinical trial challenges.*

After investigating the current clinical trial process, we have utilized the unique features of blockchain technology to have problem-oriented development for multiple healthcare applications targeting the current clinical trial challenges related to patient recruitment, patient engagement, data capture, and overall management. We have conducted a large-scale simulation on each application to evaluate its feasibility, stability, and robustness.

3. *This study has provided optimal solutions to the known blockchain limitations.*

To avoid the blockchain limitations described in Chapter 1.1, we have proposed the following solutions: (1) before operating the system, all users need to set up a username and a password which will be stored in the home healthcare facility's blockchain adapter and mapped to the public key and private key. This is a potential solution to key management vulnerability. The user can also contact the authority to rebuild a new account and remove the original one, (2) the data component in the transactions is encrypted and can only be decrypted by a certain group of users controlled by the smart contract. This will protect privacy even though the encrypted data is publicly auditable, (3) we have created a control mechanism inside each blockchain adapter as only sending five transactions every second. This setting could avoid Ethereum's scalability limitation by spacing the transactions but it does not solve the known scalability issue of Ethereum fundamentally, and (4) each blockchain node is converted to a blockchain adapter which provides the compatibility of existing tools without extensive arrangement. The design of using IPFS as complementary file storage managed by a blockchain enhances the feasibility of utilizing blockchain technology for healthcare applications.

8.4 Future work

Our future work will continue to investigate the needs of the clinical trial process and seek solutions using blockchain technology, such as adding machine learning tools to monitor patient's conditions persistently and predict side effects and overall outcomes. The current safety monitoring process described in section 6.2.4 relies on the EDC process. However, the DSMB convenes only when the clinical trial has been conducted for a while and the data has met a certain point. Adding AI components to the Study Conduct module

could achieve more efficient monitoring. We will also investigate more potential in the CTMS design using blockchain technology such as integrating secure multi-party computation with blockchain for computational applications such as subject matching and utilizing the cryptocurrency concept to build a novel clinical trial management system which will help ensure timely validation and payment.

The proposed overall blockchain design can potentially serve as a global patient-level clinical research data-sharing and analytics platform. Utilizing blockchain's unique properties, this platform could coordinate and facilitate the data sharing process while protecting patients' privacy and data security. With the cryptocurrency concept, the platform can create an incentive mechanism to honor the data of contributors and researchers.

BIBLIOGRAPHY

[1] M. Hills and P. Armitage, "The two-period cross-over clinical trial," *British journal of clinical pharmacology*, vol. 8, no. 1, p. 7, 1979.

[2] Sullivan J. Subject recruitment and retention: Barriers to success. Appl Clin Trials 2004 Apr:50-54

[3] B. Carlisle, J. Kimmelman, T. Ramsay, and N. MacKinnon, "Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials," (in eng), *Clin Trials*, vol. 12, no. 1, pp. 77-83, Feb 2015, doi: 10.1177/1740774514558307.

[4] J. P. Newhouse, "Medical care costs: how much welfare loss?," *Journal of Economic perspectives*, vol. 6, no. 3, pp. 3-21, 1992.

[5] Herson J: Data and Safety Monitoring Committees in Clinical Trials. Chapman and Hall/CRC, Boca Raton, Florida; 2009.

[6] J. S. Ancker, M. Silver, M. C. Miller, and R. Kaushal, "Consumer experience with and attitudes toward health information technology: a nationwide survey," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 152-156, 2013.

[7] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support," (in eng), *J Biomed Inform*, vol. 42, no. 2, pp. 377-81, Apr 2009, doi: 10.1016/j.jbi.2008.08.010.

[8] J. R. Vest and L. D. Gamm, "Health information exchange: persistent challenges and new strategies," *Journal of the American Medical Informatics Association,* vol. 17, no. 3, pp. 288-294, 2010.

[9] S. Funning, A. Grahnén, K. Eriksson, and 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," *The Quality Assurance Journal: The Quality Assurance Journal for Pharmaceutical, Health and Environmental Professionals,* vol. 12, no. 1, pp. 3-7, 2009.

[10] Thangaraj V, Reddy S, inventors; Reddy Somashekar N, assignee. Clincal trial management. United States patent application US 10/296,827. 2003 Nov 6.

[11] Briegs KL, Detoro D, Keim A, Saillot JL, inventors; Merck Sharp, Dohme Corp, assignee. Clinical trial management system. United States patent US 7,054,823. 2006 May 30.

[12] Dagalur S. CTMS: What you should know. Applied Clinical Trials website. [Online]. Available: http://www. appliedclinicaltrialsonline. com/clinical-trialmanagement-systems-what-you-should-know. 2016 Mar 18;29.

[13] Y. Saumell, O. Torres, M. Batista, and L. Sánchez, "Validation of instruments for assessing drug safety management during the conduction of clinical trials," *International journal of health policy and management,* vol. 7, no. 7, p. 623, 2018.

[14] Veeva, "Unified Clinical Operations Survey," 2019. [Online]. Available: https://www.veeva.com/unified-clinical-survey/

[15] Y. Zhuang, L. Sheets, Z. Shae, J. J. P. Tsai, and C. R. Shyu, "Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials," (in eng), *AMIA Annu Symp Proc*, vol. 2018, pp. 1167-1175, 2018.

[16] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Manubot; 2019 Nov 20.

[17] R. Grinberg, "Bitcoin: An innovative alternative digital currency," *Hastings Sci. & Tech. LJ*, vol. 4, p. 159, 2012.

[18] D. Shrier, W. Wu, and A. Pentland, "Blockchain & infrastructure (identity, data security)," *Massachusetts Institute of Technology-Connection Science*, vol. 1, no. 3, pp. 1-19, 2016.

[19] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on ethereum smart contracts (sok)," in *International Conference on Principles of Security and Trust*, 2017: Springer, pp. 164-186.

[20] G. Zyskind and O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops*, 2015: IEEE, pp. 180-184.

[21] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1-32, 2014.

[22] H. Zhao, P. Bai, Y. Peng, and R. Xu, "Efficient key management scheme for health blockchain," *CAAI Transactions on Intelligence Technology*, vol. 3, no. 2, pp. 114-118, 2018.

[23] McGhin T, Choo KK, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. Journal of Network and Computer Applications. 2019 Jun 1;135:62-75.

[24] T.-T. Kuo, H.-E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications," *Journal of the American Medical Informatics Association,* vol. 24, no. 6, pp. 1211-1220, 2017.

[25] L. A. Linn and M. B. Koo, "Blockchain for health data and its potential use in health it and health care related research," in *ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, Maryland, United States: ONC/NIST*, 2016, pp. 1-10.

[26] T.-T. Kuo, H. Zavaleta Rojas, and L. Ohno-Machado, "Comparison of blockchain platforms: a systematic review and healthcare examples," *Journal of the American Medical Informatics Association,* vol. 26, no. 5, pp. 462-478, 2019.

[27] Y. Zhuang, L. Sheets, Z. Shae, J. J. P. Tsai, and C.-R. Shyu, "Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials," (in eng), *AMIA Annu Symp Proc,* vol. 2018, pp. 1167-1175, 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30815159

[28] D. Shrier, W. Wu, and A. Pentland, "Blockchain & infrastructure (identity, data security)," *MIT Connection Science,* pp. 1-18, 2016.

[29] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper,* vol. 151, pp. 1-32, 2014.

[30] G. Greenspan, "Private blockchains are more than just shared databases", accessed 2015, [online] Available: http://www.multichain.com/blog/2015/10/private-blockchains-shared-databases/.

[31] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access*, vol. 4, pp. 2292-2303, 2016.

[32] Szabo N. Smart contracts: building blocks for digital markets. EXTROPY: The *Journal of Transhumanist Thought*,(16). 1996;18(2).

[33] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016: ACM, pp. 254-269.

[34] K. Delmolino, M. Arnett, A. Kosba, A. Miller, and E. Shi, "A programmer's guide to ethereum and serpent," *URL: https://mc2-umd. github. io/ethereumlab/docs/serpent_tutorial. pdf.(2015).(Accessed May 06, 2016)*, 2015.

[35] Y. Zhuang, L. R. Sheets, Y. W. Chen, Z. Y. Shae, J. J. P. Tsai, and C. R. Shyu, "A Patient-Centric Health Information Exchange Framework Using Blockchain Technology," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2169-2176, 2020, doi: 10.1109/JBHI.2020.2993072.

[36] Y. Zhuang, Y.-W. Chen, Z.-Y. Shae, and C.-R. Shyu, "Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation," *J Med Internet Res*, vol. 22, no. 7, p. e19029, 2020/7/27 2020, doi: 10.2196/19029.

[37] Pramanik MI, Lau RY, Demirkan H, Azad MA. Smart health: Big data enabled health paradigm within smart cities. Expert Systems with Applications. 2017 Nov 30;87:370-83.

[38] C. Weng *et al.*, "Using EHRs to integrate research with patient care: promises and challenges," *Journal of the American Medical Informatics Association,* vol. 19, no. 5, pp. 684-687, 2012.

[39] A. K. Jha, "Meaningful use of electronic health records: the road ahead," *Jama,* vol. 304, no. 15, pp. 1709-1710, 2010.

[40] J. Orloff *et al.*, "The future of drug development: advancing clinical trial design," *Nature Reviews Drug Discovery,* vol. 8, no. 12, pp. 949-957, 2009.

[41] P. J. Embi, A. Jain, J. Clark, S. Bizjack, R. Hornung, and C. M. Harris, "Effect of a clinical trial alert system on physician participation in trial recruitment," *Archives of internal medicine,* vol. 165, no. 19, pp. 2272-2277, 2005.

[42] L. Dimitropoulos, V. Patel, S. Scheffler, and S. Posnack, "Public attitudes toward health information exchange: perceived benefits and concerns," *The American journal of managed care,* vol. 17, no. 12 Spec No., pp. SP111-6, 2011.

[43] J. S. Shapiro, F. Mostashari, G. Hripcsak, N. Soulakis, and G. Kuperman, "Using health information exchange to improve public health," *American journal of public health,* vol. 101, no. 4, pp. 616-623, 2011.

[44] L. M. Fernandes, M. O'Connor, and V. Weaver, "Big data, bigger outcomes," *Journal of AHIMA,* vol. 83, no. 10, pp. 38-43, 2012.

[45] D. C. Mohr, M. N. Burns, S. M. Schueller, G. Clarke, and M. Klinkman, "Behavioral intervention technologies: evidence review and recommendations for

future research in mental health," *General hospital psychiatry*, vol. 35, no. 4, pp. 332-338, 2013.

[46] Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. The American Journal of Human Genetics. 2009 Nov 13;85(5):643-54.

[47] Unertl KM, Johnson KB, Lorenzi NM. Health information exchange technology on the front lines of healthcare: workflow factors and patterns of use. Journal of the American Medical Informatics Association. 2012 May 1;19(3):392-400.

[48] Buterin V. A next-generation smart contract and decentralized application platform. white paper. 2014 Jan 14;3(37).

[49] J. L. Fernández-Alemán, I. C. Señor, P. Á. O. Lozoya, and A. Toval, "Security and privacy in electronic health records: A systematic literature review," *Journal of biomedical informatics*, vol. 46, no. 3, pp. 541-562, 2013.

[50] P. Ray and J. Wimalasiri, "The need for technical solutions for maintaining the privacy of EHR," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006: IEEE, pp. 4686-4689.

[51] Y. Zhuang, L. Sheets, A. Atkins, C.-R. Shyu, Z. Shae, and C. Hsu, "A Hybrid Blockchain Design for Patient Recruitments and Persistent Monitoring for Clinical Trials," *ieee.blockchain.org*, 2018.

[52] P. Sylim, F. Liu, A. Marcelo, and P. Fontelo, "Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention," *JMIR research protocols*, vol. 7, no. 9, p. e10163, 2018.

[53] S. Treweek *et al.*, "Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis," *BMJ open*, vol. 3, no. 2, p. e002360, 2013.

[54] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: A literature review," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, 2017: IEEE, pp. 763-768.

[55] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, "Blockchain technology: Beyond bitcoin," *Applied Innovation*, vol. 2, no. 6-10, p. 71, 2016.

[56] G. Hripcsak *et al.*, "Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 204-211, 2014.

[57] W. J. Gordon and C. Catalini, "Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability," *Computational and structural biotechnology journal*, vol. 16, pp. 224-230, 2018.

[58] D. Bender and K. Sartipi, "HL7 FHIR: An Agile and RESTful approach to healthcare information exchange," in *Proceedings of the 26th IEEE international symposium on computer-based medical systems*, 2013: IEEE, pp. 326-331.

[59] S. Gummadi, N. Housri, T. A. Zimmers, and L. G. Koniaris, "Electronic Medical Record: A Balancing Act of Patient Safety Privacy Health Care Delivery," *The American journal of the medical sciences*, vol. 348, no. 3, pp. 238-243, 2014.

[60] Y. Zhuang, L. Sheets, Z. Shae, Y.-W. Chen, J. Tsai, and C.-R. Shyu, "Applying Blockchain Technology to Enhance Clinical Trial Recruitment," in *AMIA 2019 Annual Symposium*, 2019, pp. 1276-1285.

[61] C. A. Pedersen, P. J. Schneider, and J. P. Santell, "ASHP national survey of pharmacy practice in hospital settings: prescribing and transcribing—2001," *American Journal of Health-System Pharmacy,* vol. 58, no. 23, pp. 2251-2266, 2001.

[62] A. M. Heekin, J. Kontor, H. C. Sax, M. S. Keller, A. Wellington, and S. Weingarten, "Choosing Wisely clinical decision support adherence and associated inpatient outcomes," *The American journal of managed care,* vol. 24, no. 8, p. 361, 2018.

[63] N. Menachemi, S. Rahurkar, C. A. Harle, and J. R. Vest, "The benefits of health information exchange: an updated systematic review," *Journal of the American Medical Informatics Association,* vol. 25, no. 9, pp. 1259-1265, 2018.

[64] D. Blumenthal, "Stimulating the adoption of health information technology," *New England journal of medicine,* vol. 360, no. 15, pp. 1477-1479, 2009.

[65] S. Rahurkar, J. R. Vest, and N. Menachemi, "Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care," *Health affairs,* vol. 34, no. 3, pp. 477-483, 2015.

[66] K. S. Williams and S. J. Grannis, "Examining the Heartland Region Pilot: First Look at the Patient-Centered Data HomeTM Framework," in *AMIA*, 2018.

[67] (2016). *Health Information Exchange: Opportunities and Challenges for Health Centers*. [Online] Available: https://c.ymcdn.com/sites/indianapca.site-ym.com/resource/resmgr/IQIN_HCCN/HIE_data_pre_read.pdf

[68] R. S. Rudin, A. Motala, C. L. Goldzweig, and P. G. Shekelle, "Usage and effect of health information exchange: a systematic review," *Annals of internal medicine*, vol. 161, no. 11, pp. 803-811, 2014.

[69] K.-Y. Wen, G. Kreps, F. Zhu, and S. Miller, "Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey," *Journal of medical Internet research*, vol. 12, no. 4, p. e73, 2010.

[70] C. Williams, F. Mostashari, K. Mertz, E. Hogin, and P. Atwal, "From the Office of the National Coordinator: the strategy for advancing the exchange of health information," *Health affairs*, vol. 31, no. 3, pp. 527-536, 2012.

[71] J. J. Cimino, M. E. Frisse, J. Halamka, L. Sweeney, and W. Yasnoff, "Consumer-mediated health information exchanges: The 2012 ACMI debate," *Journal of biomedical informatics*, vol. 48, pp. 5-15, 2014.

[72] D. B. McCarthy, K. Propp, A. Cohen, R. Sabharwal, A. A. Schachter, and A. L. Rein, "Learning from health information exchange technical architecture and implementation in seven beacon communities," *EGEMS*, vol. 2, no. 1, 2014.

[73] L. Kolkman and B. Brown, "The Health Information Exchange Formation Guide: The Authoritative Guide for Planning and Forming an HIE in your State, Region or Community," 2011: HiMSS.

[74] H. Wu and E. M. LaRue, "Linking the health data system in the US: challenges to the benefits," *International journal of nursing sciences*, vol. 4, no. 4, pp. 410-417, 2017.

[75] Feied C, Iskandar F, inventors; Microsoft Corp, assignee. Master patient index. United States patent application US 11/683,799. 2007 Nov 8.

[76] Patil HK, Seshadri R. Big data security and privacy issues in healthcare. In2014 IEEE international congress on big data 2014 Jun 27 (pp. 762-765). IEEE.

[77] Rouse WB, Serban N. Understanding and managing the complexity of healthcare. MIT Press; 2014 Jul 3.

[78] M. Terry, "Medical identity theft and telemedicine security," *Telemedicine and e-Health*, vol. 15, no. 10, pp. 928-933, 2009.

[79] S. Simon, J. S. Evans, A. Benjamin, D. Delano, and D. Bates, "Patients' attitudes toward electronic health information exchange: qualitative study," *Journal of medical Internet research*, vol. 11, no. 3, p. e30, 2009.

[80] P. Fontaine, S. E. Ross, T. Zink, and L. M. Schilling, "Systematic review of health information exchange in primary care practices," *The Journal of the American Board of Family Medicine*, vol. 23, no. 5, pp. 655-670, 2010.

[81] S. Romanosky, R. Telang, and A. Acquisti, "Do data breach disclosure laws reduce identity theft?," *Journal of Policy Analysis and Management*, vol. 30, no. 2, pp. 256-286, 2011.

[82] M. N. Ngafeeson, "Healthcare information systems opportunities and challenges," in *Encyclopedia of Information Science and Technology, Third Edition*: IGI Global, 2015, pp. 3387-3395.

[83] P. Ranade-Kharkar, S. E. Pollock, D. K. Mann, and S. N. Thornton, "Improving clinical data integrity by using data adjudication techniques for data received through a Health Information Exchange (HIE)," in *AMIA Annual Symposium*

Proceedings, 2014, vol. 2014: American Medical Informatics Association, p. 1894.

[84] Y. Zhuang, L. R. Sheets, Z. Shae, Y.-W. Chen, J. J. P. Tsai, and C.-R. Shyu, "Applying Blockchain Technology to Enhance Clinical Trial Recruitment," (in eng), *AMIA Annu Symp Proc*, vol. 2019, pp. 1276-1285, 2020.

[85] Mahon E, Roberts J, Furlong P, Uhlenbrauck G, Bull J. Barriers to clinical trial recruitment and possible solutions: a stakeholder survey. Applied Clinical Trials. 2015 Sep;24.

[86] J. M. Watson and D. J. J. B. m. r. m. Torgerson, "Increasing recruitment to randomised trials: a review of randomised controlled trials," vol. 6, no. 1, p. 34, 2006.

[87] Marcellus L. Are we missing anything? Pursuing research on attrition. Canadian Journal of Nursing Research Archive. 2004 Sep 1:82-98.

[88] B. Carlisle, J. Kimmelman, T. Ramsay, and N. J. C. T. MacKinnon, "Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials," vol. 12, no. 1, pp. 77-83, 2015.

[89] S. Treweek *et al.*, "Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis," vol. 3, no. 2, p. e002360, 2013.

[90] R. B. Gul and P. A. J. J. o. c. n. Ali, "Clinical trials: the challenge of recruitment and retention of participants," vol. 19, no. 1-2, pp. 227-233, 2010.

[91] G. D. Huang, J. Bull, K. J. McKee, E. Mahon, B. Harper, and J. N. J. C. c. t. Roberts, "Clinical trials recruitment planning: A proposed framework from the Clinical Trials Transformation Initiative," vol. 66, pp. 74-79, 2018.

[92] Frank G. Current challenges in clinical trial patient recruitment and enrollment. SoCRA Source. 2004 Feb;2(February):30-8.

[93] G. R. Sadler, H. C. Lee, R. S. H. Lim, J. J. N. Fullerton, and h. sciences, "Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy," vol. 12, no. 3, pp. 369-374, 2010.

[94] J. K. Payne and C. C. J. A. N. R. Hendrix, "Clinical trial recruitment challenges with older adults with cancer," vol. 23, no. 4, pp. 233-237, 2010.

[95] M. X. Patel, V. Doku, and L. Tennakoon, "Challenges in recruitment of research participants," *Advances in Psychiatric Treatment,* vol. 9, no. 3, pp. 229-238, 2003, doi: 10.1192/apt.9.3.229.

[96] Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Medical care. 2002 Aug 1:IV3-18.

[97] Hufford M, Peterson D, Paty JA, Shiffman S, inventors; invivodata (reg) Inc, assignee. System for clinical trial subject compliance. United States patent US 8,065,180. 2011 Nov 22.

[98] Bleicher PA, Stamos N, Klofft JP, Dale RM, inventors; Phase Forward Inc, assignee. Clinical trial data management system and method. United States patent US 6,820,235. 2004 Nov 16.

[99] M. Benchoufi and P. Ravaud, "Blockchain technology for improving clinical research quality," *Trials*, vol. 18, no. 1, p. 335, 2017/07/19 2017, doi: 10.1186/s13063-017-2035-z.

[100] A. Baliga, I. Subhod, P. Kamat, and S. Chatterjee, "Performance evaluation of the quorum blockchain platform," *arXiv preprint arXiv:1809.03421*, 2018.

[101] C. Cachin and M. Vukolić, "Blockchain consensus protocols in the wild," *arXiv preprint arXiv:1707.01873*, 2017.

[102] A.-W. Chan *et al.*, "SPIRIT 2013 statement: defining standard protocol items for clinical trials," *Annals of internal medicine*, vol. 158, no. 3, pp. 200-207, 2013.

[103] Y. Chen, H. Li, K. Li, and J. Zhang, "An improved P2P file system scheme based on IPFS and Blockchain," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017: IEEE, pp. 2652-2657.

[104] E. Nyaletey, R. M. Parizi, Q. Zhang, and K. R. Choo, "BlockIPFS - Blockchain-Enabled Interplanetary File System for Forensic and Trusted Data Traceability," in *2019 IEEE International Conference on Blockchain (Blockchain)*, 14-17 July 2019 2019, pp. 18-25, doi: 10.1109/Blockchain.2019.00012.

[105] J. T. Mitchel and J. B. Hays, "System and method for electronic document management, organization, collaboration, and submission in clinical trials," ed: Google Patents, 2010.

[106] Roy K. Electronic Trial Master Files. *Applied Clinical Trials*. 2009 Mar 1:16.

[107] J. Viega, M. Messier, and P. Chandra, *Network security with openSSL: cryptography for secure communications*. " O'Reilly Media, Inc.", 2002.

[108] L. Mbuagbaw, L. Thabane, P. Ongolo-Zogo, and T. Lang, "The challenges and opportunities of conducting a clinical trial in a low resource setting: the case of the Cameroon mobile phone SMS (CAMPS) trial, an investigator initiated trial," (in eng), *Trials,* vol. 12, p. 145, Jun 9 2011, doi: 10.1186/1745-6215-12-145.

[109] F. Koenig *et al.*, "Sharing clinical trial data on patient level: opportunities and challenges," *Biometrical Journal,* vol. 57, no. 1, pp. 8-26, 2015.

[110] K. I. Howard, M. S. Krause, and J. T. Vessey, "Analysis of clinical trial data: The problem of outcome overlap," *Psychotherapy: Theory, Research, Practice, Training,* vol. 31, no. 2, p. 302, 1994.

[111] C. R. Breitkopf, M. Loza, K. Vincent, T. Moench, L. R. Stanberry, and S. L. Rosenthal, "Perceptions of reimbursement for clinical trial participation," (in eng), *J Empir Res Hum Res Ethics,* vol. 6, no. 3, pp. 31-38, 2011, doi: 10.1525/jer.2011.6.3.31.

[112] C. Grady, "Payment of clinical research subjects," *The Journal of clinical investigation,* vol. 115, no. 7, pp. 1681-1687, 2005.

[113] J. Benet, "Ipfs-content addressed, versioned, p2p file system," *arXiv preprint arXiv:1407.3561,* 2014.

[114] Y Zhuang, L. Sheets, X Gao, Y Shen, Z Shae, J Tsai, C-R Shyu, "Development of A Blockchain Framework for Virtual Clinical Trials," in *AMIA Annual Symposium*, 2020.

[115] G. Frank, "Current challenges in clinical trial patient recruitment and enrollment," *SoCRA Source,* vol. 2, no. February, pp. 30-8, 2004.

[116] B. Patrick-Lake, "Patient engagement in clinical trials: The Clinical Trials Transformation Initiative's leadership from theory to practical implementation," *Clinical Trials*, vol. 15, no. 1_suppl, pp. 19-22, 2018.

[117] R. B. Giffin, Y. Lebovitz, and R. A. English, *Transforming clinical research in the United States: challenges and opportunities: workshop summary*. National Academies Press, 2010.

[118] Oliveira, Pedro, H. Zejnilovic, and Helena Canhao. "Challenges and opportunities in developing and sharing solutions by patients and caregivers: the story of a knowledge commons for the patient innovation project." *Governing Medical Knowledge Commons* 301 (2017).

[119] E. National Academies of Sciences *et al.*, "The National Academies Collection: Reports funded by National Institutes of Health," in *Virtual Clinical Trials: Challenges and Opportunities: Proceedings of a Workshop*, C. Shore, E. Khandekar, and J. Alper Eds. Washington (DC): National Academies Press (US), 2019.

[120] M. Gold *et al.*, "Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 4, pp. 234-242, 2018.

[121] Stergiopoulos, Stella, et al. "Measuring the impact of patient engagement and patient centricity in clinical research and development." *Therapeutic innovation & regulatory science* 54.1 (2020): 103-116.

[122] A. Sertkaya, H. H. Wong, A. Jessup, and T. Beleche, "Key cost drivers of pharmaceutical clinical trials in the United States," (in eng), *Clin Trials,* vol. 13, no. 2, pp. 117-26, Apr 2016, doi: 10.1177/1740774515625964.

[123] D. Covington and K. Veley, "The remote patient-centered approach in clinical research," *Applied Clinical Trials,* vol. 24, no. 2/3, p. 30, 2015.

[124] R. C. Holt and J. R. Cordy, "The Turing programming language," *Communications of the ACM,* vol. 31, no. 12, pp. 1410-1424, 1988.

[125] Y. Zhuang, L. Sheets, Z. Shae, Y.-W. Chen, J. Tsai, and C.-R. Shyu, "Applying Blockchain Technology to Enhance Clinical Trial Recruitment," in *AMIA Annual Symposium Proceedings*, 2019: American Medical Informatics Association, pp. 1276-1285.

[126] Z. Lu and J. Su, "Clinical data management: Current status, challenges, and future directions from industry perspectives," *Open Access J Clin Trials,* vol. 2, pp. 93-105, 2010.

[127] L. B. Mokkink *et al.*, "The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study," *Quality of life research,* vol. 19, no. 4, pp. 539-549, 2010.

[128] A. Coulter and J. Ellins, "Effectiveness of strategies for informing, educating, and involving patients," *Bmj,* vol. 335, no. 7609, pp. 24-27, 2007.

[129] A. Coulter, *Engaging patients in healthcare*. McGraw-Hill Education (UK), 2011.

[130] R. McCorkle *et al.*, "Self-management: Enabling and empowering patients living with cancer as a chronic illness," *CA: a cancer journal for clinicians,* vol. 61, no. 1, pp. 50-62, 2011.

[131] M.-P. Pomey, D. P. Ghadiri, P. Karazivan, N. Fernandez, and N. Clavel, "Patients as partners: a qualitative study of patients' engagement in their health care," (in eng), *PLoS One,* vol. 10, no. 4, pp. e0122499-e0122499, 2015, doi: 10.1371/journal.pone.0122499.

[132] L. Singler, P. McAdams, G. Uhlenbrauck, K. Jernigan, and J. Schulman, "Models of engagement: patients as partners in clinical research," *Applied Clinical Trials,* vol. 27, no. 6, pp. 28-31, 2018.

[133] M. Maltoni *et al.*, "Prognostic factors in advanced cancer patients: evidence-based clinical recommendations—a study by the Steering Committee of the European Association for Palliative Care," DOI: 10.1200/JCO.2005.06.866 *Journal of Clinical Oncology 23, no. 25,* September 01, 2005, pp. 6240-6248.

[134] R. L. Fleurence, L. H. Curtis, R. M. Califf, R. Platt, J. V. Selby, and J. S. Brown, "Launching PCORnet, a national patient-centered clinical research network," *Journal of the American Medical Informatics Association,* vol. 21, no. 4, pp. 578-582, 2014.

[135] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Sci Data*, vol. 3, p. 160035, May 24 2016, doi: 10.1038/sdata.2016.35.

VITA

Yan Zhuang is a doctoral candidate in Health Informatics at the Institute for Data Science and Informatics. He received his B.S. degree in Electrical Engineering with honor from the University of Missouri and in Communication Engineering from the University of Electronic Science and Technology of China in 2016.

Yan's research interest focuses on the performance and sustainability of blockchain technology and utilizing blockchain's unique features integrated with AI components such as machine learning and federated learning to provide practical solutions to diverse healthcare challenges in health information exchange, clinical trial recruitment, persistent trial monitoring, electronic data capture, and overall clinical trial management. Since 2016, he started his Ph.D. training as a graduate research assistant. He has participated in NEJM SPRINT Data Analysis Challenge and won second place in the IEEE Blockchain for Clinical Trials White Board Challenge. He served as the president of the Institute for Data Science and Informatics Graduate Student Organization from 2019 to 2020. He will join the National Institute of Health Data Science at Peking University as a postdoctoral fellow in July 2021.

During his Ph.D. study, Yan has 5 first author, 1 co-first author, and 1 co-author publications as well as a few poster presentations.