

# Applying Blockchain Technology to Enhance Clinical Trial Recruitment

Yan Zhuang, BS<sup>1</sup>, Lincoln R. Sheets, MD, PhD<sup>2</sup>, Zonyin Shae, PhD<sup>3,4</sup>, Yin-Wu Chen, PhD<sup>3</sup>, Jeffrey J. P. Tsai, PhD<sup>3</sup>, Chi-Ren Shyu, PhD<sup>1,2,3</sup>

1 Informatics Institute; 2 School of Medicine, University of Missouri, Columbia, MO, USA; 3 Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan; 4 Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan

## Abstract

*Patient recruitment for clinical trials is known to be a challenging aspect of clinical research. There are multiple competing concerns from the sponsor, patient and principal investigator's perspectives resulting in most clinical trials not meeting recruitment requirements on time. Conducting under-enrolled clinical trials affects the power of conclusive results or causes premature trial termination. The Blockchain is a distributed ledger technology originally applied in the financial sector. Its features as a peer-to-peer system with publicly audited transactions, data security, and patient privacy are a good fit for the needs of clinical trials recruitment. The "Smart Contract" is a programmable self-executing protocol that regulates the blockchain transactions. Given current recruitment challenges, we have proposed a blockchain model 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.*

## Introduction

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<sup>1, 2</sup>. 86% of clinical trials don't achieve their recruitment goals on time<sup>3</sup> and 19% of registered clinical trials were either closed or terminated due to failure to reach expected enrollment<sup>4</sup>. Barriers persist although there have been many research papers addressing the challenges of identifying and recruiting subjects to clinical trials over the past decades<sup>5-9</sup>.

Barriers to recruiting patients into clinical trials can be classified into three different categories (Table 1) based on (1) sponsor perspectives, (2) principal investigator perspectives, and (3) subject perspectives<sup>3, 8</sup>. Sponsors initially need adequate participants for the potential trial to file an application with the Food and Drug Administration (FDA) for approval<sup>3</sup>. 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<sup>10</sup>. 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<sup>8, 11</sup>. Sponsors need to settle on trial sites without knowing the geographical distribution of future subjects<sup>7</sup>, but distant trial sites will deter many potential subjects<sup>8</sup>. 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<sup>4, 7, 11</sup>. 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<sup>8, 11, 12</sup>. With these persistent challenges, an efficient model is needed to enhance the recruitment process.

Table 1. Current recruitment barriers from different perspectives

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

Blockchain is a distributed ledger technology first applied in the financial sector<sup>13</sup>. 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<sup>14</sup>. Blockchain also has other features such as decentralization,

immutability, ensuring data provenance, and public auditability<sup>14</sup>. All the transactions that have occurred in the blockchain are distributively stored into each node which can be any active electronic device without hardware requirements inside the blockchain system. Any transaction needs to be validated by the users in the blockchain before it is written into the system<sup>15</sup>. Since the system is fully decentralized, all the transactions can be audited publicly by all the users. The public auditability feature can solve the issue of lack of awareness.

The Smart Contract is a self-executing, coded protocol agreed between senders and receivers initially proposed by Nick Szabo in 1996 to regulate all transactions on the distributed ledger system<sup>16</sup>. Most blockchain systems have added a smart contract function to their protocols such as Ethereum and Hyperledger<sup>17</sup>. Since the Ethereum blockchain comes with a built-in Turing complete programming language used for the smart contract, any computational problem such as patient matching for recruitment, or checking the validity of a clinical trial can be coded as a smart contract<sup>18</sup>. Once a smart contract is deployed into the system, the system will generate an application binary interface (ABI) of the smart contract and smart contract address for users to call the smart contract functions<sup>14</sup>. All the users need to follow the smart contract's regulation to make a transaction. The ABI and address are viewable to all the users rather than the source code or the data stored inside the smart contract, but the functions may not be executable to any user. This depends on the user's privilege. For example, in the clinical trial recruitment setting, the inclusion/exclusion protocol can only be input into a smart contract by the authority who is in charge of all clinical trials. This ideally would be the FDA. Only the sponsors can send a matching request to the smart contract, then the blockchain system can automatically match the potential subjects. This matching of sponsors and subjects can help the sponsor to target eligible patients, save time for principal investigators and patients to understand the protocols, and ensure the completeness of the protocol. Once the eligible subjects are identified, the sponsor can select the trial sites based on the subjects' geographical distribution which could solve the issue of inaccessible clinical sites for subjects. The smart contract can also ensure the validity of a clinical trial by checking the sponsor's identity and whether the trial has a National Clinical Trial (NCT) identifier number which means the study has been approved by the authority<sup>19</sup>. Patients can rest assured that the trial is legal, and that the sponsors are reliable. Original blockchain features and smart contract functionality can solve most of the barriers for clinical trial recruitment.

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<sup>20</sup>; inability to retrieve any information once a user has lost the private key, (2) loss of privacy caused by transparency of the distributed system<sup>21, 22</sup>; ability of all the users in the blockchain to view all the data stored in the blockchain, and (3) scalability constraints<sup>22, 23</sup> considering Ethereum can handle roughly 15 transactions per second. Most blockchain applications in the healthcare area are still in the design stage and have not yet been implemented<sup>24, 25</sup>. The abovementioned challenges are mainly caused by blockchain's features instead of actual operation. We have implemented a blockchain system using Ethereum blockchain which is an open source platform with a built-in smart contract function to tackle the issues for clinical trial recruitment. We have tested our system by simulating clinical trial recruitment using a real dataset to test the feasibility and provide potential solutions to current common blockchain challenges. We have made the following assumptions to perform the simulation of the recruitment process using our system: (1) the authority, each clinical site and each sponsor needs to provide at least one node in the system; (2) patients have to opt-in to our system, they authorized the system to access their health records; (3) clinical sites agreed to connect the secured Electronic Health Records (EHR) database through blockchain node; (4) all the EHR are recorded in the same standard; (5) there is a universal patient ID to map the same patient across different clinical sites; (6) the authority, each clinical site and each sponsor needs an administrator to operate the system, patients need to operate the system as well.

## **Blockchain**

Blockchain is a fully distributed peer-to-peer network<sup>14</sup>. The blockchain system runs based on users' consent rather than being managed by a third party. To protect the users' privacy, every user in the blockchain will have a unique key pair which contains a public key and a private key to represent their identities<sup>26</sup>. The public key is similar to a user's bank account. The private key is similar to a user's signature. Every transaction needs to be digitally signed by the sender's private key<sup>26</sup>. Once a user makes a transaction, all the users in the blockchain can see the sender's public key instead of their real identity. The blockchain has a built-in mechanism to check whether the sender's public key and the signed private key match, but keeps the private key hidden from other users<sup>13</sup>. The stability of the blockchain relies on all nodes to provide computing power to validate the transactions. Once a transaction which is one-time exchange of data is made, the transaction will be sent to each user in the blockchain for validation. If a hacker wants to hack others' accounts to made transactions using a fake private key, the transaction will be voted down by other users and discarded. Where blockchain was originally used in the financial sector, the transaction referred to an exchange

of cryptocurrency like bitcoin. The validation process validates whether the sender has sufficient balance to make the transaction. In our system, a transaction refers to the execution of smart contract function such as authority input inclusion/exclusion criteria of a certain trial into a smart contract. All the users will validate whether the sender's public key is the authority, the data which is criteria will be encrypted and can only be decrypted by specific users instead of all users. Blockchain is a chain of blocks<sup>13</sup>. Each block contains a unique block number, last block number and all of the transactions that occurred after the last block was generated. The block numbers chain the blocks together that can be used for tracing the source of the data<sup>13</sup>. For instance, the sponsors need to check the patients' records to check the eligibility for the trial, they can check whether the records were input by the hospitals by tracing the sender of the initial transaction. Once a transaction is verified and written into the block, the record and data cannot be altered anymore.

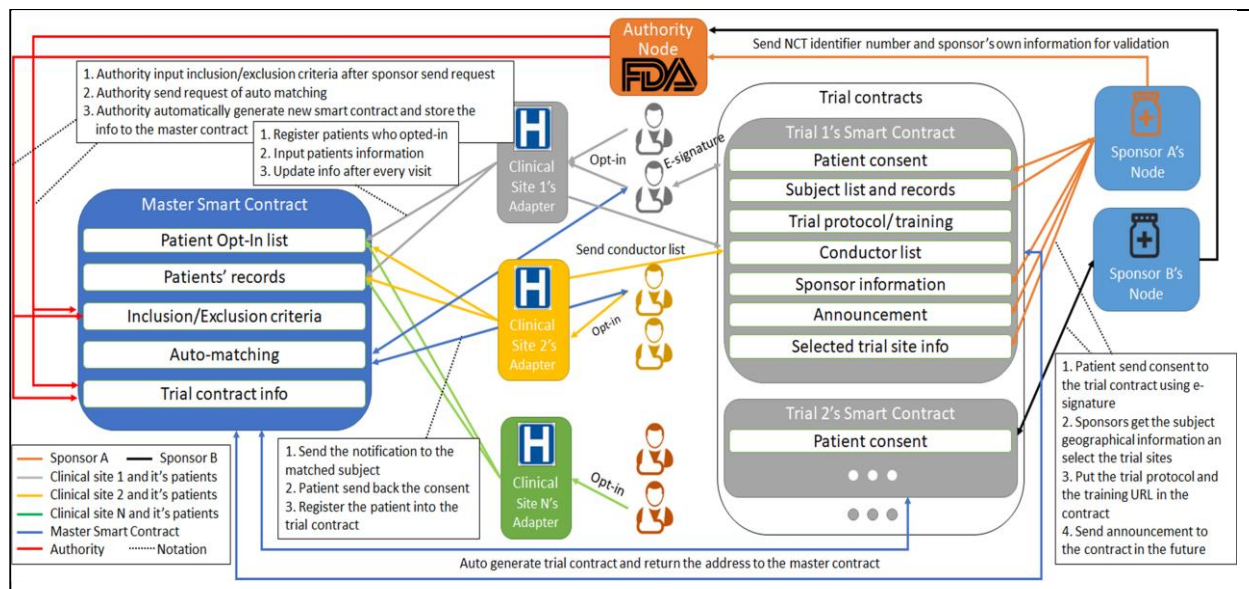
Blockchain can be built as a "public chain" or a "private chain"<sup>14</sup>. The public chain is also known as a "permissionless chain" meaning all the users can join the public chain without additional setups and permission. "Bitcoin" and "Ether" are applications on the public blockchain and public Ethereum blockchain, respectively. The private chain is called the "permission chain". The creator of the private chain has control over who can join the chain. In order to join the private chain, all the nodes need to deploy the same "genesis block" which is the starting block of the private chain provided by the creator and add at least one peer node which already in the chain. The "genesis block" of a private chain determines the private chain's characteristics such as the ability to estimate block generation rate. The genesis blocks are different from different blockchains so that they can only be acquired from the creators. In our scenario, the authorities will be the creator of the system. All the nodes provided by the clinical sites and sponsors need to get the "genesis block" file from the authority. The "genesis block" file is a JSON file that can be embedded into a program used for blockchain system installation. All the nodes also need to add the authority's node as a peer using its IP address and node identifier which is automatically generated by the blockchain. This procedure can also be embedded into the installment program provided by the authority. Since the blockchain is a fully distributed system, the creator can decide who has permission to join the chain. Other nodes have the same privileges to operate the system such as sending transactions, using smart contracts, auditing transactions, etc.

Since blockchain is not allowed to communicate with servers outside of the chain<sup>27</sup>, we need to rely on a blockchain adapter to check the patient's EHR data and push the result back to the blockchain in order to precisely match the subject with complex inclusion and exclusion criteria. For example, the sponsors from different IPs send requests to a clinical site's adapter to identify potential trial candidates from EHR systems. Each node provided by the clinical site needs to be built as a "blockchain adapter" to join our system. Blockchain adapter is a Remote Procedure Call (RPC) server which can connect the clinical site's own secured EHR database protected by the hospital's firewall. RPC server executes functions after clients from different IPs send requests to the server. When a patient opts-in to the blockchain system from a clinical site, this site's administrator needs to create a blockchain account for the patient through their node. The blockchain account will then be associated with the patient's universal patient ID in the EHR database. The administrator needs to create accounts for clinical research coordinators (CRC) as well. The authority also needs to build its node into a blockchain adapter to communicate with the NCT database in order to check whether the NCT number and the sponsor's information has been matched.

## **Implementation**

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 1) 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.

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 2). 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.



**Figure 1.** System architecture master smart contract and different clinical trial smart contracts

**Table 2.** 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
	Inclusion/Exclusion 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 into 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 2.

```

//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]);
          }
        }
      }
    }
  }
}
}

```

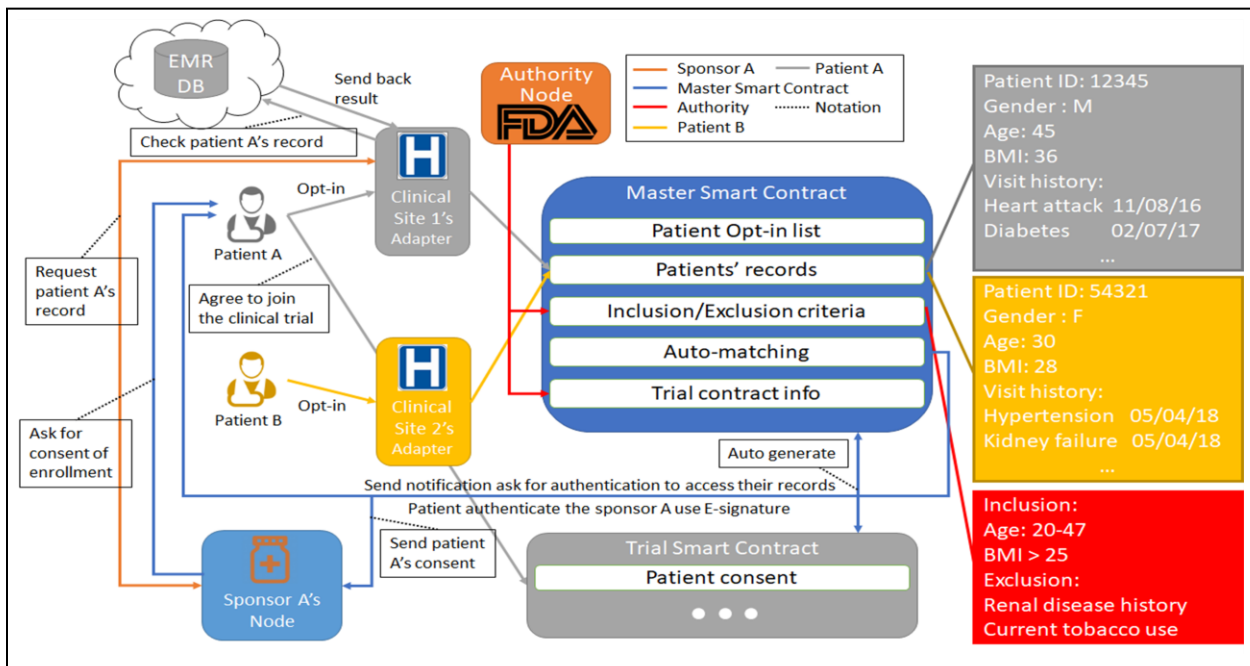
```

"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 2.** 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.)

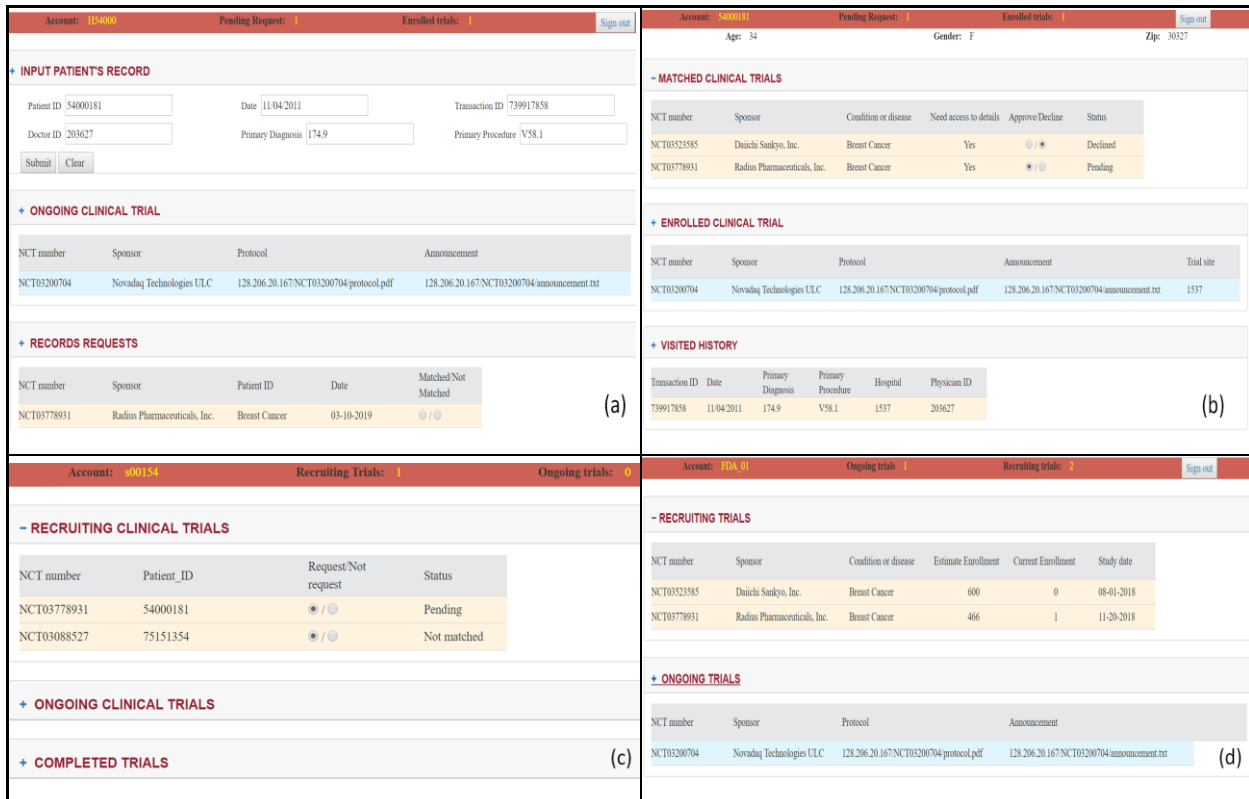
The patient matching is a two-step process. The first step is using the auto-matching function which can only narrow down the patient selection but cannot perform precise matching due to the complexity of the inclusion/exclusion criteria. The master smart contract will automatically match the criteria with the patients' records stored in the smart contract to select potential subjects. Referring to the example shown in Figure 3, the master smart contract filtered out patient B due to the exclusion criteria of history of renal disease. The second step is performed by the clinical sites to precisely match the potential subjects' health records in the hospital's secured database with the patient's information from the consents. The matching process is then followed by sending the result back to the blockchain through its own blockchain adapter. The exclusion criteria in the example in Figure 3 also has 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 3.** An example of 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 into the trial-based 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 Graphical User Interface (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 4(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 4(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 4(c)). The authority will have all the clinical trial lists and the trial details (Figure 4(d)).



**Figure 4.** 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

### 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 (NCT03778931, NCT03252431, NCT03367572, NCT02068092, NCT03152929, NCT03354611, NCT01185132, NCT03523585, NCT03529110, NCT03200704). 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 3. All the selected frequent criteria 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 Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute for the simulated cases. The SEER database contains cancer incidence and survival data covering 34.6 % of the U.S. population<sup>28</sup>. For our simulation, we randomly picked 6,000 patients with breast cancer from the SEER database and evenly distributed them into three clinical sites.

**Table 3.** 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.

## 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 address of the trial-based contract is 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

```
> master_contract.check_matched("NCT103200704")
1145
> master_contract.get_info("NCT103200704")
{"0xc8b9f2a936af717FE3F29acf08d4644c3220Caf3"
> NCT103200704_contract.address
"0xc8b9f2a936af717FE3F29acf08d4644c3220Caf3"
> NCT103200704_contract.get_trial_subjects_amount("")
100
> NCT103200704_contract.get_trial_subjects_geo_info("")
["31723|",|"32125|",|"39866|",|"39791|",|"083381",|"084311",|"34736|",|"31174|",|"37230|"]
```

**Figure 5.** Calling smart contract functions to check the trial NCT103200704's info from the sponsor's account

will use the GUIs (Figure 4(b)) 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 5 is automatically generated by the master smart contract.

## Discussion

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 is inputted by the authority and all the clinical sites share the same version of the protocol, there won't be an insufficient trial protocol. 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 each 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 at 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 user 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 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.



## Conclusion and Future work

Through our simulation process, our blockchain model shows features that can tackle the 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 as timesaving, 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. Our future work is to further evaluate Ethereum's validating mechanism which could fundamentally solve the scalability issue. Our extended study will be to add an artificial intelligence component to our system for the real-time detection of significant adverse events of the collected EHR during a clinical trial.

## Acknowledgment

This research is supported in part by the University of Missouri Informatics and Data Science Research Initiatives (YZ and CRS), MOST108-2634-F-468-001 from the Ministry of Science and Technology through Pervasive Artificial Intelligence Research (PAIR) Labs Taiwan, and MOST 106-2632-E-468-002 (ZS, YWC, and JT).

## References

1. 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;24.
2. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. *BMC medical research methodology*. 2006;6(1):34.
3. Sullivan J. Subject recruitment and retention: barriers to success, *Appl Clin Trials*, 2004.
4. Carlisle B, Kimmelman J, Ramsay T, MacKinnon NJCT. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. 2015;12(1):77-83.
5. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. 2013;3(2):e002360.
6. Gul RB, Ali PAJocn. Clinical trials: the challenge of recruitment and retention of participants. 2010;19(1 - 2):227-33.
7. Huang GD, Bull J, McKee KJ, Mahon E, Harper B, Roberts JNJcct. Clinical trials recruitment planning: A proposed framework from the Clinical Trials Transformation Initiative. 2018;66:74-9.
8. Frank G. Current challenges in clinical trial patient recruitment and enrollment. *SoCRA Source*. 2004;2(February):30-8.
9. Furlong EMJRP, Bull GUJ. Barriers to Clinical Trial Recruitment and Possible Solutions: A Stakeholder Survey. 2015.
10. Sadler GR, Lee HC, Lim RSH, Fullerton JJN, sciences h. Recruitment of hard - to - reach population subgroups via adaptations of the snowball sampling strategy. 2010;12(3):369-74.
11. Payne JK, Hendrix CCJANR. Clinical trial recruitment challenges with older adults with cancer. 2010;23(4):233-7.
12. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. *Advances in Psychiatric Treatment*. 2003;9(3):229-38.
13. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008.
14. 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;2018:1167-75.
15. Crosby M, Pattanayak P, Verma S, Kalyanaraman VJAI. Blockchain technology: Beyond bitcoin. 2016;2(6-10):71.
16. Szabo NJETJoTT. Smart contracts: building blocks for digital markets. 1996;18.
17. Christidis K, Devetsikiotis MJIA. Blockchains and smart contracts for the internet of things. 2016;4:2292-303.
18. Buterin V. A next-generation smart contract and decentralized application platform. white paper. 2014;3:37.
19. Mandatory Reporting of National Clinical Trial (NCT) Identifier 2014.
20. Zhao H, Bai P, Peng Y, Xu R. Efficient key management scheme for health blockchain. *CAAI Transactions on Intelligence Technology*. 2018;3(2):114-8.
21. McGhin T, Choo K-KR, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. *Journal of Network and Computer Applications*. 2019.
22. Kuo T-T, Kim H-E, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association*. 2017;24(6):1211-20.

23. Linn LA, Koo MB, editors. Blockchain for health data and its potential use in health it and health care related research. *ONC/NIST Use of Blockchain for Healthcare and Research Workshop Gaithersburg, Maryland, United States: ONC/NIST*; 2016.
24. Kuo T-T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. *Journal of the American Medical Informatics Association*. 2019;26(5):462-78.
25. Mackey TK, Kuo TT, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - challenges and opportunities for applications of blockchain technology in the future of healthcare. *BMC Med*. 2019;17(1):68.
26. Zyskind G, Nathan O, editors. *Decentralizing privacy: Using blockchain to protect personal data*. 2015 IEEE Security and Privacy Workshops; 2015: IEEE.
27. McConaghy T, Marques R, Müller A, De Jonghe D, McConaghy T, McMullen G, et al. *BigchainDB: a scalable blockchain database*. 2016.
28. Institute NC. Overview of the SEER Program.