**A Blockchain Portal for Clinical Trials, Preliminary White Paper**

**Author: Daniel Wong**

**Introduction**

(State need for service and current state of clinical trial research—lack of transparency, traceability, real time access, and “paper trail” audit, also data falsification—writing intro should be trivial)

**Methods**

A simulation involving a completed and deposited trial in ImmPort

Upon approval to begin phase 2, the FDA instantiates a blockchain for the clinical trial, and registers all participating parties in the portal. All parties are required to use the portal service for any exchange of information, and only the information present on the blockchain will be used for review when considering approval of the drug.

Upon agreement of the protocol by the FDA and completion of patient recruitment and consent by the clinical site, the clinical site gives a unique verification code and a private key to each patient prior to visiting the clinician.

**Proposed schema:**

For iterations in clinical trial phases 2 -> 3:

-pharma posts encrypted placebo distribution key on ledger

-transaction(sender: pharma, receiver: FDA, data: placebo key, timestamp) added to ledger

-pharma/CRO specifies how to write CRF form, company managing portal codes it as html form and posts it on portal for clinicians to fill out

-clinician receives verification code from patient during visit and inputs into portal

-clinician or clinical research staff inputs exam data through portal

-validation of CRF:

-patient that saw clinician matches patient assigned to clinician through verification code assigned to patient—presence of valid code to ensure interaction of clinician and patient

-(still under consideration): patient sent encrypted digital CRF and asked to sign for accuracy (through digital RSA signature, each patient given a private key by CRO)

-wilder ideas: cons are that patient needs to have some app on phone

- upon patient entering, clinician indicates patient present on portal, push notification sent to patient with time sensitive verification code, GPS to ensure patient is in vicinity of office for t time

-upon patient entering, clinician indicates patient present on portal, push notification sent to patient with time sensitive verification code (or barcode for clinician to scan to ensure physical presence, can detect if screen shots taken on patient phone), second push sent t time later upon patient indicating finished procedure

-patient can just be paid to wait in lobby in both scenarios

-pro is that we can measure how long patient was in office for and compare to how long procedure should take

-encrypted CRF recorded on blockchain ledger upon validation and sent to CRO: transaction(sender: clinicianj, receiver: CRO, data: {patientj, CRFij}, timestamp)

-upon recording on ledger, adverse effects and safety concerns extracted from CRF and broadcast (publicly, no encryption needed) to DSMB for viewing on the portal

-if outside data collection sources are employed, data collected (e.g. blood tests) encrypted and sent to CRO

-transaction(sender: outside data collection company, receiver: CRO, extra data, timestamp) added to ledger

-CRO decrypts, performs data analysis and clean up

-once finalized, sends encrypted data to pharma

-transaction(sender: CRO, receiver: pharma, cleaned data, timestamp)

-pharma performs stats and writes analysis scripts, transaction(sender: pharma, receiver: FDA, data: statistical analyses + scripts, timestamp) posted to ledger

-Pharma applies for FDA approval

-FDA accesses portal and views entire transaction history of phase 1 -> 3 with full read access to all data, performs validation of blockchain and files (hashing checks) and then performs audit

-If approved:

-FDA selectively makes data public for view by anyone on the portal (clinicaltrials.gov schema)

-if denied:

-repeat iteration of trial phase or kill trial

**Views and Architecture:**

Users:

-FDA

-Clinicians

-Outside data collection sources (organizations like blood test folk)

-Contract research organization (CRO)

-Data safety management board (DSMB)

-Pharma

-General Public

Views:

1. Clinician input: raw and secured html form to fill out CRF
2. FDA/admin view: blockchain ledger with all transaction histories, and all downloadable files (similar to version 1 demo), and access to DSMB page
3. Public page: page for the general public, can list adverse effects and general info on trials like clincialtrials.gov, information like safety on ongoing trials, and selectively disclosed raw information decided by FDA
4. DSMB view: display all clinical trials that the DSMB is involved with
   1. Each clinical trial will have a page that lists all participating patients and any adverse side effects reported on the CRF, all in real time as the trial is happening
5. Third party data collector view: see all clinical trials involved with, form to send data to CRO of specific trial
6. Pharma view: see all clinical trials involved with, link to ledger and safety concerns similar to DSMB
7. CRO view: form to send data to Pharma

Architecture:

-Private blockchain (access to participate in transactions is granted by FDA)

-Distributed data base with replication and duplication across different physical machines, managed by the regulator

- byzantine fault tolerance through majority vote of warehouse nodes

-SHA256 hashing checks periodically run throughout trial to ensure data integrity across all warehouse machines

-append only scheme, if edits need to be made to submitted forms (like the CRF), post a new transaction with the new data, old data maintained in the blockchain

-save CRF data as JSON data type for ease of parsing, reading, and data handling

-gather potential transactions into query set and periodically process set of potential transactions to make sure ordering is the same across all warehouse nodes

-encrypt everything with private public key?

**Results**

Can show link to demo instantiated with views and real clinical data from SDY1 trial. Give reviewer admin privileges and access to nodes in the network to use.

**New Stuff and Why This is Cool:**

-Open transparency of all transactions happening in the network

-Immutable data of any type: CRFs, statistical analysis and scripts, third party data like blood tests, cleaned data from CRO

-chronological audit trail since phase 1

-example audit: if want to validate CRO—can compare raw data from clinicians to cleaned data from CRO

-centralized portal for all things clinical trial (like clinicaltrial.gov but better with raw data, advantages of blockchain infrastructure)

-move away from corruptible paper documents and faxes, digitize and encrypt, and selectively allow people to read—a good compromise between conflict of personal privacy and need to analyze data

-zero knowledge proof of data purity:

-example with placebo distribution:

-pharma claims data was untouched, provable without actually revealing sensitive data contents

-pharma cannot modify placebo distributions to better fit study goals, and can pass audit by hashing encrypted data and comparing to ledger

-like a secure and encrypted GitHub for clinical trials

-will perhaps be effective in international trials where we are much less confident in integrity

**Discussion:**

Forcing all participating parties to use a service like this may be a challenge. This can be overcome by restricting approval of a drug upon the condition that the pharmaceutical company uses the service for information exchange, and that they hire CROs and clinicians with the same use condition. Only data on the blockchain will be considered when reviewing potential approval of a drug. Although blockchain technology provides a means of recording data into structures that are immutable, traceable, and verifiable, it cannot prevent data from being falsified at the point of origin. Clinicians can be careless or fraudulent and record misleading data into the CRFs, and statisticians within pharmaceutical companies can overinflate p-values because of vested interests in success. These mistakes are carried forward in the blockchain. However, with unfalsifiable data collected from clinicians and cleaned up by CROs before being sent to pharma, we have the raw data to validate the statistical results that come out of pharmaceutical companies, all due to the design of blockchain. Independent statistical analyses can be run to verify the results of pharma. The pharmaceutical statisticians and CROs can also be required to post their python and R analysis scripts and freeze these on the ledger for verification.

By providing each patient with a verification code to give to their physician and validate the CRF upon visit, patient interaction is ensured. Unfortunately, if a patient and fraudulent clinician are in full cooperation, there is no way of assuring that real data is generated for that office visit. In theory, physicians can fraudulently purchase these verification codes from patients to save time and avoid running the test trial, but adding a verification code provides a deterrent from blatantly fabricating CRFs. Fabricating data will require the involvement and cooperation of the patient. This will be less enticing and riskier for the clinician to generate fake CRFs, as the clinician is no longer the only party with knowledge of the forgery. Discovery of malpractice is now more likely and can make administering the office visit treatment more attractive than fabricating data. Additionally, moving clinicians away from paper based recordings and forms may also be difficult, but not impossible as the move towards EHRs rises and as the public concern for non-standardized and fallible physical documentation increases. If a service like this can be adopted, the benefits of immutability, traceability, and more trust in the clinical research process can be strengthened. Regulators can confidently track the complex flow of data throughout a trial, and the public can be kept up to date on safety proceedings and progress.