Research

A Cryptographic Method for Reproducible, Transparent Randomization of Clinical Trial Design

Corresponding Author: Dr Abdul Badih El Ariss
Co-Authors: Ian Miers, PhD, Mr David Chen, Colin McSwiggen, PhD, Ms Yan yan, Mr Sammer Marzouk, Ms Deborah Lai, Matthew Green, PhD, Mr Norawit Kijpaisalratana, Ali Raja, MD, Dr Jarone Lee

Edited By: Dr Shuhan He

CONDUCTSCIENCE
ACADEMIC PUBLISHING

Corresponding Author: Dr Abdul Badih El Ariss
Co-Authors: Ian Miers, PhD, Mr David Chen, Colin McSwiggen, PhD, Ms Yan yan, Mr Sammer Marzouk, Ms Deborah Lai, Matthew Green, PhD, Mr Norawit Kijpaisalratana, Ali Raja, MD, Dr Jarone Lee
Type: Research
Keywords: cryptographic, randomization
Edited By: Dr Shuhan He
Received: 2024-05-03
Published: -

A Cryptographic Method for Reproducible, Transparent Randomization of Clinical Trial Design

There remains difficulty in creating adequate and equitable protocols and mechanisms to guide the use of randomization to assign participants to different treatment arms in a randomized clinical trial. A trustworthy randomization system should be correct, reproducible, confidential, and transparent, to ensure fairness, secure public trust, and protect both the recruited participants and the trial supervisors. We propose the design and implementation of an accountable and trustworthy randomization approach using a verifiable pseudorandom shuffling mechanism for the purposes of simple randomization in clinical trials. The solution has been deployed in the form of a user-accessible, no-download website for ready access to the public.

Background

Complete randomization is a common and basic method used to generate similar-sized groups of participants without consideration of participant characteristics [1]. This technique eliminates the potential for selection bias and negligible likelihood of group size and co-variate imbalances in large clinical trials (n > 200) [2]. The use of the most valid randomization technique remains a critical consideration when designing a robust randomized clinical trial [3]. Recent studies have highlighted concerns in the reproducibility of randomized clinical trials due to limited reporting of study sampling and randomized participant assignment methods necessary to produce generalizable study findings [4], [5], [6].

Transparency of study protocols for randomized clinical trials is necessary to generate unbiased comparison groups, but the majority of surveyed study protocols for controlled trials fail to include

adequate information to describe their randomization methods [7]. The need for transparency in how participants are assigned in a randomized clinical trial motivates the need for the development of a verifiable, complete randomization approach to facilitate trial reproducibility and validity.

Method

Objectives

We designed a pseudorandom rather than random procedure to enable retroactive verification, which is an important feature for verification of clinical trial randomization. A *pseudorandom* process is a deterministic, and therefore repeatable, process that nonetheless produces outputs indistinguishable from random outputs [8]. For practical purposes, this means that the results of the process are verifiable, but are impossible to predict or manipulate without knowing all of the inputs in advance.

The proposed system is implemented as a webpage that takes a list of participant identifiers (or other similar values) as input and produces an ordering of the list. Key features of the system are as follows:

1. *Correctness*. The ordering of the list is cryptographically *pseudorandom*: this means that each participant has an equal probability of being at any position in the list, and the ordering cannot be predicted before the list is known. The system operates effectively with any number of participants, though as with all randomization methods, larger sample sizes (n > 200) provide better balance of covariates between groups.
2. *Retroactive verifiability*. Each lottery is verifiable after the fact. Given a list of participant identifiers used as input, and some auxiliary information such as the date, it is possible to deterministically reproduce the lottery.
3. *Confidentiality*. The system does not involve the transmission or disclosure of any confidential participant data. No participant data is ever sent from the local computer to a remote server.
4. *Transparency*. The system does not depend on any secret values, such as passwords or cryptographic keys, that could fall into the hands of an unauthorized person and be used to manipulate the outcome.
5. *Tamper resistance*. The system is resistant to attempts to manipulate the lottery outcome, for example by changing the lottery participants or their participant IDs.

Web application design

The proposed system consists of two static HTML webpages, deployed directly from an open-source Git repository at https://github.com/diracdeltas/accountable-triage via GitHub Pages. All computation happens locally via client-side JavaScript in the browser. No participant data is sent from the local browser, thereby avoiding medical privacy concerns. The two pages are as follows:

1. https://accountable-triage.cs.jhu.edu/: This provides a user interface for users to enter a list of Medical Record Numbers (MRNs) and presents the lottery outcome as a shuffling of the original list, with the MRNs ranked in order of highest priority participant to lowest priority participant.
2. https://accountable-triage.cs.jhu.edu/verify: This provides a user interface for users to enter a list of MRNs and a past date to verify the result of a past lottery. To correctly verify the time-dependent input value from NIST, it is also necessary to enter the time zone in which the randomization procedure was carried out.

Cryptographic shuffling procedure

We designed our cryptographic lottery procedure with the following technical specifications in mind: (1) each list is permuted in a pseudorandom manner, (2) this permutation can be verified after the fact, (3) the creator or operator of the system cannot bias the lottery without detection, and (4) small changes to the list of participants are unlikely to change the outcome for any given participant. The system operates on a complete list of participants and currently supports simple randomization only. Block randomization is not supported in the current implementation. The system can process both individual participants and groups, but must be run with the complete list of participants to maintain randomization integrity.

To prevent the results from being predictable before the day on which the randomization procedure is run, the input to the system includes random values published online by the Interoperable Randomness Beacons project of the National Institute of Standards and Technology (NIST) [9]. In addition, we apply the random oracle model of using hash functions to produce deterministic output that is cryptographically pseudorandom (repeatable and non-randomized) [10].

To satisfy the design objectives of the proposed randomization procedure, we first append the random beacon value to the end of each participant ID to generate one unique identifier for each participant ID. The random beacon value used is the earliest value published by the NIST Randomness Beacon after midnight on the current date in the local time zone where the lottery is being run. Second, we input each unique identifier, composed of both the participant ID and the appended random beacon value, into the random oracle (SHA3-512 hash function) to create a hash value that represents each participant's lottery number. Third, we sort the lottery numbers lexicographically and generate an ordered list. The resulting ordered list of lottery numbers, each associated with a unique participant ID, constitutes the lottery outcome. Figure 1 visualizes the cryptographic randomization procedure. In the case of healthcare triage where new lottery participants present in real time, a new randomized lottery can be run that will generate a unique, independent outcome.

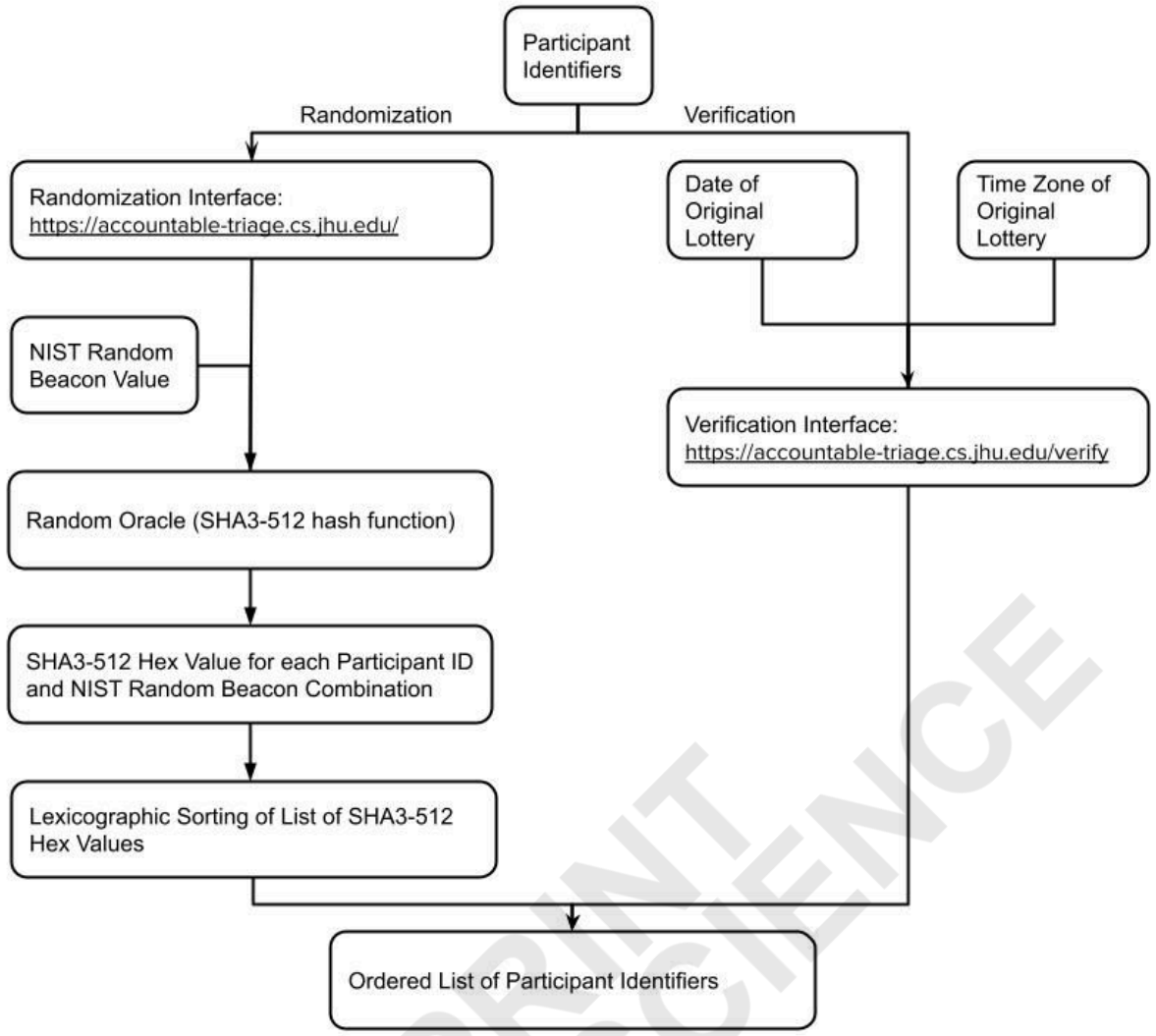

Fig. 1. Diagram of the proposed cryptographic randomization method using participant identifiers as input and verification of the original randomization result.

The current implementation is limited to simple randomization and does not support cluster randomization, multiple treatment arms, or cross-over designs. While the system can be accessed remotely from multiple sites, it does not include built-in functionality for stratification by site. Multi-site trials would need to implement additional protocols to manage site-specific randomization or stratification.

Results

To show the validity and use cases of the proposed cryptographic method for randomization of scarce healthcare resources and clinical trial design, we conducted a reproducible case study using synthetic data. The objective of this validation step for the case study for randomized clinical trial design is to 1) successfully randomize a set of synthetic identifiers that represent participants using the proposed cryptographic method and 2) generate a reproducible result that can be independently verified post-randomization.

We entered 10 synthetic participant identifiers on a lottery date of February 5, 2023 and time zone of GMT-5:00 (https://accountable-triage.cs.jhu.edu/), see the Triage sheet in Supplementary Table 1 for participant identifiers. The lottery result was generated (https://accountable-triage.cs.jhu.edu/verify), see Supplementary File 1 for the ordered list and associated date and time zone metadata. We successfully verified the results of the original lottery by inputting the same

10 synthetic participant identifiers for verification at a later date (February 12, 2023) along with the original lottery date of February 5, 2023 and time zone of GMT-5:00 and produced the same ordered list as the original lottery.

Discussion

Our objective is to provide scientists with an approach for random participant assignment in clinical trials that offers correctness, reproducibility, confidentiality, and transparency.

One limitation of the system stems from the time interval between changes in the time-dependent input value. Each day at midnight, the relative ranking of participant IDs for the next 24-hour period becomes available to users of the randomization system. The system will reproduce the same results in perpetuity as long as the NIST beacon data remains accessible, allowing for long-term verification of randomization outcomes. This could create a vulnerability if it were possible to change participants' participant IDs on the day of the randomization, after the ranking for that day has been determined. We therefore recommend that the participant IDs used for the lottery be assigned according to a definite process that is not subject to human discretion.

A related limitation is the possibility of running two randomization lotteries with the same participant IDs in two different 24-hour periods, enabling one to choose between the two outcomes. We assume however that this type of use will rarely be feasible, due to the time-sensitive nature of triage decisions. There are alternative cryptographic solutions to this limitation, but they impose design trade-offs such as requiring a remote server that extend beyond the scope of the local design of our method.

We also do not address the possibility of users entering incorrect records and changing priorities for group assignments outside of the randomization algorithm. Our system does not prevent such overt manipulations. However, the system ensures that such actions are visible and traceable, after the fact. We suggest that it be made clear to users of this system that all assignments of clinical trial participants can be verified.

Conclusion

We have designed a cryptographic randomization system that is deterministic, time-dependent, and verifiable to ensure public trust in randomized assignment of participants in clinical trials. The system uses cryptographic methods, takes the form of two static web pages that can be hosted locally on a hospital computer without transmitting any participant data to a remote server, and is deployed directly from an open-source repository visible to the public. Prospective users of this system should take its limitations into account and ensure that proper protocols are in place so that it can be used securely and as intended.

Supporting information

Supplementary File 1. Synthetic participant identifiers used as input for cryptographic randomization as part of the two validation case studies.

Data Availability

The code and data underlying this study can be found at the Accountable Triage GitHub repository: https://github.com/diracdeltas/accountable-triage.

List of abbreviations

CSC: Crisis standards of care

ICU: Intensive care unit

MRN: Medical record number

NIST: National Institute of Standards and Technology

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The open source code used in this study is available in the GitHub repository, https://github.com/diracdeltas/accountable-triage.

Competing Interests

The authors declare no competing interests.

Funding

The research of Colin McSwiggen is partially supported by the National Science Foundation under Grant No. DMS 1714187. Matthew Green was funded by the National Science Foundation under awards CNS-1653110 and CNS-1801479 and the Office of Naval Research under contract N00014-19-1-2292, as well as a Security and Privacy research award from Google.

Guarantor

The guarantor is SH.

Contributorship

IM, DC, and CM researched literature, conceived the study, developed the software, and wrote the manuscript. YZ developed the software and edited the manuscript. SM, DL, and MG edited the manuscript. AR, JL, and SH designed the analysis, edited the manuscript, and supervised the team. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Acknowledgements

No acknowledgements are reported.

References

1. Suresh KP. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. Journal of Human Reproductive Sciences. 2011;4:8.
2. Lachin JM. Properties of simple randomization in clinical trials. Controlled Clinical Trials. 1988;9:312–26.

3. Berger VW, Bour LJ, Carter K, Chipman JJ, Everett CC, Heussen N, et al. A roadmap to using randomization in clinical trials. BMC Medical Research Methodology. 2021;21.
4. Riley SP, Swanson BT, Brismée J-M, Sawyer SF, Dyer EJ. Low reproducibility of randomized clinical trials methodology related to sampling: A systematic methodological review. Journal of Manual & Manipulative Therapy. 2019;27:258–66.
5. Wang SV, Sreedhara SK, Schneeweiss S, Franklin JM, Gagne JJ, Huybrechts KF, et al. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nature Communications. 2022;13:5126.
6. Vinkers CH, Lamberink HJ, Tijdink JK, Heus P, Bouter L, Glasziou P, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. PLOS Biology. 2021;19.
7. Lai D, Wang D, McGillivray M, Baajour S, Raja AS, He S. Assessing the quality of randomization methods in randomized control trials. Healthcare. 2021;9(4):100570.
8. Ripley BD. Thoughts on pseudorandom number generators. Journal of Computational and Applied Mathematics. 1990;31:153–63.
9. NIST Computer Security Resource Center: Interoperable Randomness Beacons. 2022. https://csrc.nist.gov/projects/interoperable-randomness-beacons. Accessed 19 Dec 2022.
10. Canetti R, Goldreich O, Halevi S. The random oracle methodology, revisited. Journal of the ACM. 2004;51:557–94.