Privacy-Preserving Multi-Center Clinical Analytics with Threshold Homomorphic Encryption: A Pediatric Craniospinal Irradiation Use Case

George C. Turcaş[1,2], George Gugulea[1], Cristian Lupaşcu[1], Mihai Togan[1,3], Andrada Crina Turcaş[2,4*]

[1]certSign R&I, Bucharest, Romania.
[2]Babeş-Bolyai University, Cluj-Napoca, Romania.
[3]Military Technical Academy Ferdinand I, Bucharest, Romania.
[4]The Oncology Institute Prof. Dr. Ion Chiricuţă, Cluj-Napoca, Romania.

*Corresponding author(s). E-mail(s): andrada.turcas@ubbcluj.ro;

Abstract

Purpose: Multi-center studies are often necessary in rare pediatric conditions, but direct pooling of patient-level data is constrained by privacy, governance, and hospital-integration barriers. We examine whether routine clinical statistics can be reproduced over encrypted data for a fixed hospital consortium.

Methods: We implemented a threshold version of the Cheon–Kim–Kim–Song (CKKS) homomorphic encryption scheme using OpenFHE, an open-source fully homomorphic encryption library, for Pearson correlation, Wilcoxon rank-sum, and the chi-squared test. The clinical use case concerned hematologic toxicity in a **51**-patient pediatric craniospinal irradiation cohort. We emulated two and three participating sites by partitioning this single-center cohort and used synthetic cohorts separately for scaling tests. We also describe a proposed governance model covering key-share custody, query approval, auditable decryption, and hospital-side extraction, transformation, and loading (ETL) and encryption.

Results: Across the tested cohort sizes, the encrypted chi-squared routine remained within **0.5**% of the plaintext baseline. Pearson correlation stayed within approximately **0**–**5.7**% relative error, including about **5**% at **512** patients after parameter tuning. Wilcoxon was the fragile case: relative error in the reported $\boldsymbol{Z}$ statistic was **0%** at $\boldsymbol{n = 4}$ and approximately **13.5**–**21.9%** for $\boldsymbol{n = 8}$–**64**, with no successful results reported at larger tested sizes.

Conclusion: Threshold homomorphic encryption can already support a narrow but relevant class of cross-site clinical statistics without moving plaintext patient records. The main caveats are fragile encrypted ranking for Wilcoxon and the need for explicit decryption governance when client-aided steps are used. The contribution is therefore an applied workflow, architecture, and governance model for a fixed hospital consortium rather than a production-ready platform or a new cryptographic primitive.

Keywords: homomorphic encryption, multi-center clinical analytics, clinical data governance, privacy-preserving analytics

1 Introduction

Multi-center analysis is often unavoidable in rare-disease and pediatric oncology research because no single institution has enough patients to support stable statistical inference. The same setting that creates scientific value also creates privacy risk: small cohorts, detailed longitudinal variables, and treatment-specific features make simple de-identification an unreliable protection on its own [1, 2]. The problem is not limited to direct identifiers. Dates, dose summaries, and unusual clinical trajectories can all become linkage points when data are pooled across sites. This tension has been repeatedly identified as a barrier to rare-disease collaboration and to broader health-data reuse [3].

Several privacy-preserving approaches address parts of this problem. Privacy-preserving record linkage can help sites determine whether records refer to the same patient without disclosing identifiers [4]. Differential privacy can support selected release scenarios, but the privacy-utility trade-off becomes difficult in small datasets and for statistical analyses that rely on rare or extreme outcomes [5]. Trusted execution environments and secure multiparty computation reduce some of the operational burden in other settings, but they come with different trust, interaction, and deployment assumptions [6, 7]. For the present use case, we focus on homomorphic encryption (HE), because it allows a computation service to evaluate predetermined statistics on encrypted patient-level data while keeping raw values locally, at the data owner, outside the computing server visibility [8]. Recent work indicates that this direction is moving beyond proofs of concept toward clinically recognisable analyses on real multi-institution data. In oncology-focused collaborations, multiparty FHE has been used to support survival analysis, logistic regression, and descriptive statistics without exposing raw patient records during computation [9]. In parallel, multicenter survival modelling via Cox regression has also been demonstrated under homomorphic encryption in a hospital-style setting [10]. Complementing these application systems, statistics-oriented HE toolkits are emerging that explicitly target numerical, ordinal, and categorical analyses over encrypted data [11].

This paper studies a concrete applied workflow rather than a new cryptographic construction. We revisit a pediatric craniospinal irradiation study on hematologic toxicity [12] and implement three routine statistics over encrypted data: Pearson correlation, the Wilcoxon rank-sum test, and the χ^2 test. The workflow uses threshold CKKS as implemented in OpenFHE [13, 14], so that no single party can decrypt outputs unilaterally. Performing non-polynomial operations (such as divisions) on ciphertexts using homomorphic encryption is extremely difficult and one usually has to approximate them with polynomial approximations. This is not always possible, so the final division and square-root standardization steps remain the practical bottlenecks, performed using a client-aided phase within this work. The paper also makes explicit the operational governance needed around the client-aided phase that releases approved outputs.

The target deployment is a fixed hospital consortium with a shared feature schema, a narrow hospital-side integration boundary, and explicit governance around approved output release; We study am applied HE workflow for this consortion, not a new HE scheme or cryptographic primitive.

The contribution is fourfold:

- a threshold-CKKS workflow and architecture for reproducing three routine clinical statistics over encrypted multi-center data in a fixed consortium;
- a governance and integration-boundary model for hospital consortia covering key-share custody, query approval, output release, audit, failure handling, and local hospital-side responsibilities;
- a quantitative evaluation of core encrypted evaluation runtime and numerical fidelity on a real pediatric use case, consortium-emulated cohort partitions, and synthetic scale-out cohorts;
- practical guidance on where the workflow is already usable, where the architecture remains fragile, and why governance is part of the technical contribution rather than an afterthought.

2 Related Work

Work on HE in medicine has moved from early demonstrations of encrypted biomedical analytics toward more complete application pipelines. Early studies showed that predictive analysis and selected statistical tasks could be evaluated over encrypted medical data [15], while later work

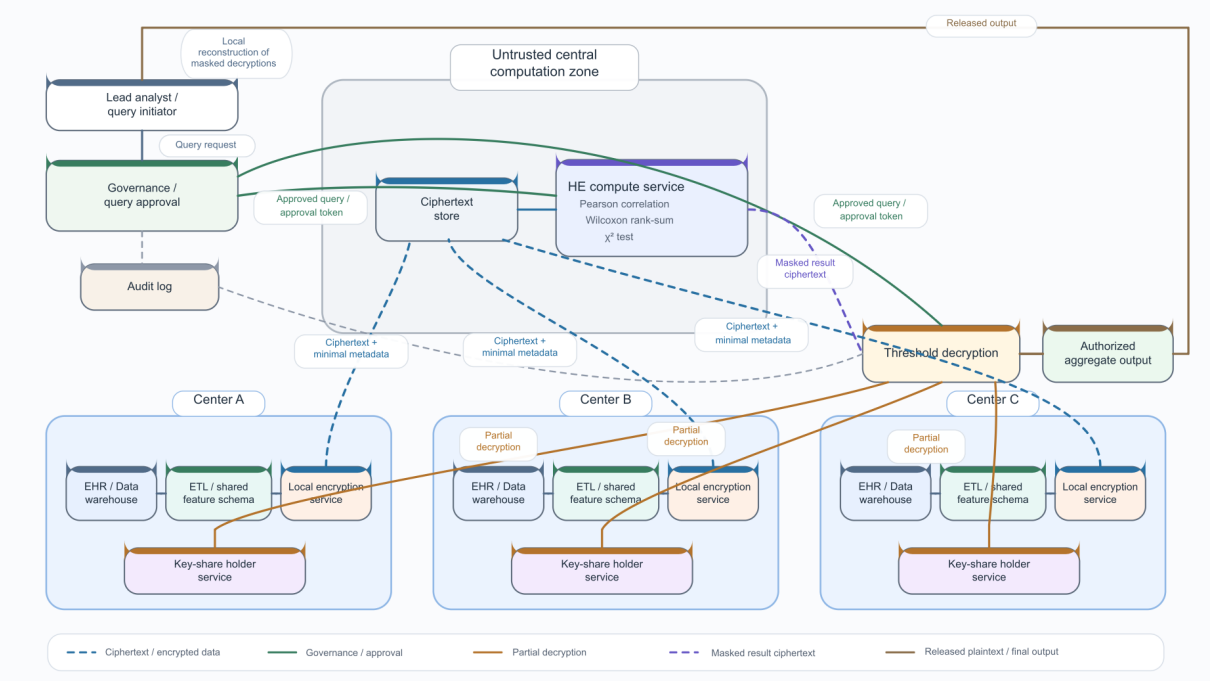

Fig. 1 Threshold-HE workflow for fixed-consortium analytics: local ETL/encryption at each center, ciphertext evaluation at an untrusted server, and approved threshold decryption of masked outputs. One-time consortium setup, including distributed key generation, joint public-key establishment, key-share provisioning, and public evaluation-key distribution, is omitted for clarity; the figure shows the per-query workflow.

explored real-time learning and EHR-oriented sharing models [16, 17]. More recent reviews document a rapidly expanding HE literature in healthcare, but also emphasize that computational cost, key management, and workflow integration remain persistent barriers to routine use [8]. Our paper fits this line of work, but narrows the scope to a small set of routinely interpreted clinical statistics and a single pediatric radiotherapy use case.

The paper is also close to recent work on multiparty or threshold HE for collaborative biomedical analysis. Lu et al. described a threshold-HE architecture for privacy-preserving multicenter logistic regression [18]. Froelicher et al. showed that multiparty HE can support federated analytics in precision medicine without centralizing plaintext data [19]. These systems demonstrate that collective decryption can be made practical, but they focus on different analysis targets and do not center the hospital-governance questions raised by a smaller, query-oriented clinical statistics workflow. More recently, multiparty FHE has been applied to collaborative oncological analytics with an explicit emphasis on practical clinical endpoints, including survival analysis, logistic regression training, and descriptive statistics [9]. For encrypted statistical analysis, HEaaN-STAT provides a broader statistical toolkit for large-scale numerical, ordinal, and categorical variables under approximate HE, illustrating the trend toward reusable statistical components rather than single-function demonstrations [11]. These works are complementary to our focus: we restrict the analysis surface to three routinely interpreted test statistics and emphasise a threshold-HE deployment model and decryption governance for a fixed hospital consortium. Unlike prior systems, we focus on three routine statistics and their release governance.

At the scheme level, threshold HE and multi-key HE offer different trade-offs. Threshold CKKS distributes one secret key across a fixed group of participants [20, 21], while multi-key HE preserves each contributor's own key and can support later onboarding more naturally [22]. We use threshold CKKS because the target setting is a fixed consortium executing approved statistical queries over a shared feature schema. The trade-off is

that onboarding or offboarding a center typically requires a fresh setup, so deployment feasibility depends on governance discipline and key lifecycle procedures, not only on ciphertext arithmetic.

For non-polynomial operations, our workflow depends on two adjacent lines of work. Cheon et al. provide an efficient homomorphic comparison method under CKKS [23], and Mazzone et al. extend this direction with ranking and order-statistics primitives that are particularly relevant to Wilcoxon-style routines [24]. Client-aided or hybrid HE systems have likewise used controlled decryption or scheme-switching phases to handle operations that are expensive to keep entirely under HE [25, 26]. We build on these ideas and our contribution consists of showing how these ingredients behave in a concrete clinical workflow and what operational controls are needed when such steps are present.

Alternative privacy-enhancing technologies remain important comparators. Secure multiparty computation can support exact collaborative statistics with different trust and interaction costs, and has recently been deployed in a European pilot study with real patient data and extensive governance documentation [7]. Differential privacy is valuable when the task is controlled release rather than exact joint computation, but current health-research evidence still shows limited routine adoption and persistent utility concerns in smaller datasets [5]. Broader PET reviews further note that TEEs can offer stronger practical performance for some workloads, at the cost of relying on hardware trust and side-channel mitigations [6]. Our work consists of the combination between a real pediatric use case, threshold HE implementation, explicit governance, and measured statistical fidelity for a small set of routine tests.

3 Methods

3.1 Clinical use case and data

Hematologic toxicity is a frequent and clinically relevant complication in pediatric patients undergoing craniospinal irradiation because large volumes of active bone marrow are exposed during treatment. This can lead to severe lymphopenia or neutropenia, increase infection or bleeding risk, and in some cases disrupt treatment delivery. We use the clinical cohort reported in [12], which contains data from 51 pediatric patients, as the real-data basis for the encrypted replication.

To emulate a multi-center setting, we partitioned the cohort into two and three site-specific subsets and treated these partitions as distinct institutional contributions. The goal was not to create a new biomedical claim from repartitioned data, but to evaluate whether a hospital consortium could recover the same routine statistics without transferring plaintext patient records. Plaintext data never left the originating institution's secure environment; only encrypted feature values participated in the cross-site workflow.

The dataset contains continuous variables, such as radiation dose and post-treatment blood counts, and categorical variables, such as toxicity indicators and treatment technique. These support three statistical routines that are clinically interpretable in this use case: Pearson correlation for continuous dose-toxicity associations, Wilcoxon rank-sum for dose comparisons across toxicity groups, and χ^2 for categorical associations. Table 1 summarizes the analyses reproduced in the encrypted setting.

This matches the analysis family reported in the source clinical study: Wilcoxon rank-sum for quantitative comparisons and chi-square for qualitative associations; the encrypted paper focuses on reproducing the corresponding test statistics inside the workflow.

To evaluate computational scaling, we also generated synthetic cohorts up to 512 patients that preserve the variable structure and medical units of the original dataset but were used only for performance and numerical-stability testing. They were not used to support new biomedical conclusions. The scientific interpretation remains supported by the original clinical study [12], while the synthetic cohorts are a systems-evaluation device.

Rare-disease studies often operate under severe sample-size constraints, which is one reason multi-center pooling matters in the first place [27, 28]. The workflow described here is meant for exactly the type of collaboration in which each individual site does not have enough data on its own.

3.2 Threshold HE and system model

We use the CKKS homomorphic encryption scheme because the target statistics involve real-valued clinical variables and mixed arithmetic over sums, products, and normalized quantities

Table 1 Clinical statistics reproduced in the encrypted workflow.

Statistical test	Variables analyzed
Pearson correlation	Dmean Pelvic Bones (Gy) vs. absolute lymphocyte count one week after radiotherapy Dmean Cervical Spine (Gy) vs. absolute neutrophil count one week after radiotherapy
Wilcoxon rank-sum	Pelvic bone dose: Grade ≥ 3 lymphopenia vs. lower grades Cervical spine dose: Grade ≥ 3 neutropenia vs. lower grades
χ^2	Radiation technique (3DCRT vs. VMAT/HT) vs. occurrence of Grade 4 lymphopenia

[13]. CKKS supports approximate arithmetic on packed vectors and is implemented in OpenFHE [14], which we use for the present prototype.

The workflow relies on threshold HE rather than multi-key HE. In a threshold setting, the consortium jointly creates one public key for encryption and splits the corresponding secret key across participating institutions. No single institution can decrypt results on its own; a predefined threshold of key-share holders must contribute partial decryptions. This model fits a fixed consortium that is willing to agree on a shared setup and on explicit approval rules for decryption. The trade-off is reduced flexibility for dynamic onboarding, where multi-key HE is often more natural [22]. Because the present paper studies a fixed hospital consortium, threshold HE is the more practical choice.

Our threat model is semi-honest. The computation service and participating institutions follow the protocol but may try to infer additional information from what they legitimately observe. Communication channels are assumed to be authenticated and encrypted. We do not claim protection against side-channel attacks, endpoint compromise, or denial-of-service. Public metadata include the number of samples, HE parameters, dataset and key version identifiers, and any other protocol information that does not itself reveal patient-level features.

The computation service receives encrypted inputs and public evaluation material, performs the homomorphic circuit, and never sees plaintext patient records. For operations that remain impractical to complete entirely under HE, most importantly the final division and square-root step after the homomorphic aggregation or ranking stage, the workflow uses a client-aided phase. In that phase the service releases only masked intermediate quantities for approved threshold decryption, after which an authorized analyst reconstructs the intended final statistic. The cryptographic workflow therefore cannot be separated from operational governance; the latter determines when a decryption is permissible, who may request it, and what is actually released.

3.3 Operational model and governance

The cryptographic design is only one part of a deployable multi-center workflow. In practice, confidentiality also depends on how keys are held, how queries are approved, what metadata are attached to each job, and how institutions recover from ordinary operational failures. This subsection makes those controls explicit.

3.3.1 Roles and trust boundaries

We model five operational roles in the consortium.

Data contributor/center. Each participating center extracts approved variables from its local EHR or data warehouse, maps them to the consortium feature schema, and encrypts the resulting vectors locally under the joint public key. Raw patient data are not released. The center releases only ciphertexts and minimal metadata (such as number of patients) needed to execute the approved workflow.

Key-share holder. Each center designates an institution-controlled service or function that stores one threshold key share and participates in approved partial decryptions. The key share is under institutional, not personal, control. A staff change therefore does not imply a change in cryptographic ownership if the service identity and institutional authorization remain intact.

Lead analyst/query initiator. The lead analyst prepares an approved analysis request, identifies the dataset and cohort version to be used, and receives the final released outputs once decryption has been authorized and completed. This role

is permitted to see the intended statistics, not raw site-level or patient-level data.

Computation service. The computation service is the untrusted server that stores ciphertext inputs, applies the homomorphic circuit, maintains job state, and produces masked outputs for the client-aided phase when required. It sees ciphertexts, public evaluation keys, public metadata (such as number of patients), and job-control information, but not plaintext clinical values.

Governance and audit function. The consortium governance function approves query types, manages decryption policy, records audit trails, and enforces rate or budget limits on decryptions. It is not a cryptographic role by itself; it is the institutional process that turns high-level policy into concrete authorization artifacts checked by the system.

3.3.2 Key management lifecycle

Distributed key generation. The consortium initializes a threshold CKKS key pair using a distributed key-generation procedure. The result is one joint public key for encryption and one secret-key share for each authorized key-share holder. No party ever reconstructs the full secret key. In the present implementation, we instantiate t-of-t configurations with $t \in \{2, 3\}$.

Authenticated setup. The setup assumes authenticated participants. Operationally, each key-share holder should authenticate using institution-managed credentials, and the resulting key identifier should be recorded together with the participating centers, threshold value, approved dataset family, and activation date.

Storage of key shares. The prototype demonstrates threshold decryption, but it does not implement a full production key-management stack. For deployment, key shares should be stored in institution-controlled services with hardened access control, ideally backed by HSM or enclave-supported secret storage. This is a deployment recommendation, not an implemented feature.

Evaluation keys and public computation material. Rotation, relinearization, and other evaluation keys are treated as public computation material associated with a specific key identifier and parameter set. They can be distributed to the computation service without enabling decryption.

Dataset binding, rotation, retirement, and consortium changes. Every ciphertext upload should be bound to a dataset identifier, cohort version, and key version. Key rotation should be triggered when policy changes, a share is suspected compromised, or consortium membership changes materially. Offboarding a center requires retiring the affected key version and re-encrypting future uploads under a fresh setup. Onboarding a new center is therefore possible, but not lightweight: it generally requires a new consortium key and a new approved data collection cycle. The threshold choice itself remains a governance trade-off between availability and collusion resistance.

3.3.3 Query approval, access control, and output release

In the intended deployment, only approved queries trigger decryption. Each request should be identified by a unique **Query ID** and should be bound to a specific dataset or cohort version, key version, approved statistic, requesting role, and validity period. Governance approval would result in a signed or otherwise verifiable **Query Approval Token**, which the computation service and key-share holder services could verify before participating in output release.

Before decryption, the intended authorization check should confirm at least four items: the request originates from an approved role, the Query Approval Token is valid and unexpired, the query budget for the relevant user or study has not been exhausted, and the requested output matches the pre-approved output type for that query. This matters because the client-aided phase is where masked intermediates become visible to the authorized analyst.

The released material must be restricted to the intended outputs only: the final statistic, plus any explicitly public quantities derived from public parameters. Raw intermediates, per-center aggregates, and reusable masked tuples should not be released as a matter of convenience. The system should log the Query ID, dataset version, key version, approver identity or service, requesting analyst, timestamp, and released output type for every successful decryption.

3.3.4 Failure handling and operational robustness

Operational failures are expected and should not force ad-hoc workarounds. Each analysis run should therefore carry a durable **Job ID** and operate on immutable encrypted inputs. If the computation service crashes after ciphertext upload but before output release, the job can be safely restarted because the inputs are ciphertext artifacts tied to a specific dataset and key version. Re-running the job should reproduce the same encrypted pipeline, subject to the randomness implicit in the encryption process and the masking randomness used in the final approved release phase.

If one key-share holder is temporarily offline, the impact depends on the threshold choice. In the current t-of-t experiments, output release waits until all required shares are available. In a production deployment, a lower threshold could improve availability, but only at the cost of weaker collusion resistance. This is a consortium policy decision that must be fixed at setup time.

If the operator responsible for a key-share service changes while the institution remains in the consortium, the key share should remain attached to the institutional service identity, not to the individual person. If a key share is suspected compromised, the active key version must be retired, the incident logged, and future analyses migrated to a fresh setup. Historic ciphertexts may also need re-encryption, depending on retention policy and threat assessment.

3.3.5 Integration with hospital IT systems and data pipeline boundary

The intended integration boundary is deliberately narrow. Each center performs local extraction from its EHR, data warehouse, or research database; maps the approved fields to a fixed consortium feature schema; and encrypts the resulting vectors before any external transmission. Only ciphertexts and minimal metadata leave the institutional environment.

The minimal metadata are limited to what the threat model already treats as public, such as dataset size, feature identifiers, key version, HE parameter set, and job-control identifiers. Communications between centers, governance endpoints, and the computation service should run over authenticated and encrypted channels. This boundary keeps the local ETL burden visible in the architecture, clarifies what the untrusted server receives, and directly motivates the workflow shown in Fig. 1.

3.4 Secure evaluation of the statistical functions

The secure routines are intended to reproduce the same test-statistic family used in the source clinical analysis, while making the workflow boundary explicit about which steps remain homomorphic and which final standardization steps require approved client-aided release.

3.4.1 Pearson correlation coefficient

We compute Pearson correlation on encrypted vectors of clinical measurements. Let n be the public cohort size. For each patient $i \in \{1, \dots, n\}$, the encrypted values x_i and y_i represent, for example, the mean dose to the pelvic bones and the absolute lymphocyte count one week after radiotherapy. For exposition, we describe a two-center partition; the same workflow extends directly to the three-center setting used elsewhere in the manuscript.

In plaintext, the statistic is

$$\begin{aligned} r &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}, \\ \overline{x} &= \frac{1}{n}\sum_i x_i, \qquad \overline{y} = \frac{1}{n}\sum_i y_i. \end{aligned} \tag{1}$$

Square roots and divisions are the expensive part under HE, so the computation service performs the homomorphic part first:

1. It centers the encrypted data and obtains ciphertext vectors $\vec{X}$ and $\vec{Y}$.
2. It computes

$$\begin{aligned} \mathsf{XdotY} &= \sum_i X_i Y_i, \\ \mathsf{LenX} &= \sum_i X_i^2, \\ \mathsf{LenY} &= \sum_i Y_i^2. \end{aligned} \tag{2}$$

3. To avoid direct encrypted division, it samples non-zero masks u_1 and u_2, sets $u_3 = u_1/u_2$, and forms the encrypted tuple

$$\big(u_1\,\mathsf{XdotY},\ u_2^2\,\mathsf{LenX},\ u_3^2\,\mathsf{LenY}\big).$$

4. The masked tuple is released only through an approved threshold-decryption step. The authorized analyst then reconstructs

$$\frac{u_1}{u_2 u_3}\,\frac{\sum_i X_i Y_i}{\sqrt{\sum_i X_i^2}\sqrt{\sum_i Y_i^2}} = r$$

without learning the unmasked components individually.

The masking step is central to the privacy analysis in Section 3.5. In brief, the client-aided phase is designed to reveal the final statistic while reducing direct exposure of the underlying numerator and norm terms. Governance controls remain necessary because repeated approved decryptions can still accumulate leakage over time.

3.4.2 Wilcoxon rank-sum test

We implement Wilcoxon on encrypted clinical values grouped by a binary toxicity indicator. Because the comparison primitive in [23] and the ranking construction in [24] operate over bounded numeric ranges, the continuous input values are first normalized to $[0, 1]$ using plaintext minimum and maximum values supplied as part of the approved analysis configuration:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}.$$

Encrypted ranking then proceeds through a compare-and-count approach. After homomorphic replication of the input vector into row-wise and column-wise encodings, pairwise comparisons yield encrypted approximate ordering bits b_{ij}, from which ranks are assembled as

$$R_i = 1 + \sum_{j \neq i} b_{ij}, \qquad b_{ij} \approx \mathbf{1}(x_i' > x_j').$$

Let $g_i \in \{0, 1\}$ denote the encrypted group label, with $g_i = 1$ for the higher-toxicity group. The encrypted rank sum and group sizes are

$$S_1 = \sum_{i=1}^{n} R_i g_i, \qquad n_1 = \sum_{i=1}^{n} g_i, \qquad n_0 = n - n_1.$$

The Wilcoxon U statistic and standardized Z statistic follow as

$$U_1 = S_1 - \frac{n_1(n_1+1)}{2}, \qquad Z = \frac{U_1 - \frac{n_0 n_1}{2}}{\sqrt{\frac{n_0 n_1 (n_0+n_1+1)}{12}}}.$$

The current prototype reports the large-sample Wilcoxon rank-sum Z statistic computed from encrypted ranks; we do not claim exact reproduction of tie-corrected or continuity-corrected p-values as reported by standard statistical software. Comparisons and ranking remain homomorphic, while only the final standardization uses a brief client-aided masked step, similar to the one for Pearson: the server masks the numerator and denominator terms, the parties threshold-decrypt them, and the client reconstructs Z in the clear. The main practical limitation is not threshold decryption itself but the ranking pipeline. In our experiments, the ranking primitive consumed enough depth and precision budget that evaluation remained reliable only up to $n \leq 64$. We therefore present Wilcoxon as a proof of concept capability: results clearly away from the significance threshold are informative, but near-threshold results should be verified with a more accurate follow-up analysis.

3.4.3 The χ^2 test

For the categorical association analysis, each center encrypts binary indicator vectors for radiotherapy technique and severe lymphopenia. The computation service first forms encrypted cell counts for the 2×2 contingency table:

$$\begin{aligned}
\mathrm{Enc}(a) &= \sum_i (1 - T_i)(1 - L_i),\\
\mathrm{Enc}(b) &= \sum_i (1 - T_i) L_i,\\
\mathrm{Enc}(c) &= \sum_i T_i (1 - L_i),\\
\mathrm{Enc}(d) &= \sum_i T_i L_i.
\end{aligned}$$

Encrypted row totals, column totals, and the public total N then give the expected counts

$$\mathrm{Enc}(E_{ij}) = \frac{1}{N}\,\mathrm{Enc}(r_i)\,\mathrm{Enc}(c_j), \qquad i,j \in \{0,1\}.$$

Because N is public, the division by N is implemented as plaintext-ciphertext multiplication rather than encrypted inversion.

The statistic is

$$\mathrm{Enc}(\chi^2) = \sum_{i,j\in\{0,1\}} \frac{\big(\mathrm{Enc}(O_{ij}) - \mathrm{Enc}(E_{ij})\big)^2}{\mathrm{Enc}(E_{ij})}.$$

The present workflow reproduces the classical Pearson χ^2 statistic on the encrypted 2×2 table; it does not claim a Yates-corrected or exact-test variant unless that variant is explicitly implemented. As in the Pearson workflow, the four masked numerator-denominator pairs are released only through the approved client-aided phase, after which the authorized analyst reconstructs the four cell contributions and sums χ^2 locally. Compared with Pearson and especially Wilcoxon, this routine is the most numerically stable in our experiments because it requires less depth and avoids encrypted ranking.

3.5 Security and privacy analysis

Protected assets. The protected assets are the patient-level feature values, site-local cohorts before encryption, intermediate encrypted aggregates, and the threshold secret-key shares. The intended cleartext outputs are limited to the final approved statistics.

Adversary model and trust assumptions. We assume semi-honest participants. The computation service may inspect ciphertexts, public evaluation material, and job metadata, but it does not deviate from the protocol. Participating institutions may likewise inspect what they legitimately receive, but no coalition smaller than the decryption threshold can decrypt on its own. Authenticated and encrypted channels are assumed. Side-channel attacks, endpoint compromise, malicious code injection, and denial-of-service are outside the present claim set.

Public metadata and intentional cleartext release. Public metadata include cohort size, HE parameters, key version, dataset or cohort version, feature schema identifiers, and job-control identifiers. The intentionally released cleartext outputs are the final approved statistics and other quantities derived only from public parameters. We do not claim to hide the fact that a given approved query was executed.

Leakage in the client-aided phase. The client-aided phase intentionally reveals the final approved statistic, and it may reveal additional information through masked tuples if closely related decryptions are approved repeatedly. For the Pearson construction, with the masking distribution used in the prototype, recovering any one hidden component of the masked triple to 1% relative accuracy at 95% confidence would require on the order of 4×10^5 independent decryptions of the same form. The Wilcoxon and χ^2 routines use the same positive multiplicative masking idea for their final client-aided step. Their exact constants differ because Wilcoxon releases a smaller masked set, while χ^2 releases one masked numerator-denominator pair per cell, but the attack surface is of the same order: estimating any one hidden term to similar accuracy still requires hundreds of thousands of repeated approved decryptions of the same query form. Masking therefore does not eliminate leakage; it attenuates single-release exposure while making repeated near-duplicate approvals the realistic abuse path. In practice, the privacy boundary is enforced by query approval, expiry, dataset and key-version binding, audit logs, and explicit decryption budgets.

Threat / control / residual-limitation summary. The following operational summary is not a formal malicious-security proof.

Threat	Primary control	Residual limitation
Curious computation service	local encryption; threshold CKKS; no plaintext on server	public metadata remain visible; side channels out of scope
Sub-threshold collusion or single-site compromise	institution-controlled key shares; threshold decryption required	availability versus collusion depends on threshold; endpoint compromise out of scope
Repeated approved decryptions of masked outputs	query approval token, expiry, dataset/key-version binding, audit log, decryption budgets	governance must be enforced operationally; repeated near-duplicate approvals can still accumulate leakage
Malicious computation service or malformed masks	not prevented in the current prototype; future verifiable computation and accountable threshold decryption discussed	current claim set is semi-honest, not malicious security

Why governance matters. Because masked tuples are still meaningful cryptographic outputs, decryption governance is part of the privacy boundary. Query IDs, approval tokens, dataset binding, key-version binding, audit logs, and decryption budgets are therefore security controls, not administrative decoration. They limit the opportunity for an analyst to request repeated near-duplicate decryptions and to combine released outputs outside the intended study design.

Integrity limitation. We do not claim malicious security. A malicious computation service could in principle alter masks or deviate from the intended circuit. A malicious participant could also attempt to manipulate approval or release workflows outside the semi-honest model. Stronger guarantees would require verifiable computation, proofs about mask formation and circuit compliance, and verifiable or accountable threshold decryption. Recent work on verifiable privacy-preserving survival analysis under multi-key FHE illustrates one concrete path in this direction (in the Cox regression setting), by enabling correctness checks against a semi-malicious compute provider [29]. Those mechanisms are future work rather than current properties of this prototype.

4 Evaluation Setup

All reported measurements were executed on a MacBook Pro with Apple M2 Pro / 12 cores (8P+4E), 16 GB unified memory, and the OpenFHE-based implementation described above. We evaluated one-party HE, two-party threshold HE, and three-party threshold HE settings where applicable.

Parameterization followed circuit depth. For Pearson, configurations scaled with the number of samples, from 1+20 bits (ring dimension 8192, depth 3) to 1+25 bits (ring dimension 16384, depth 6–8). At $n = 512$, we increased the depth and scale to mitigate noise accumulation; the tuned run is the one reported in the final summary table. Wilcoxon used a 1+23-bit configuration, ring dimension 16384, and depth 10 with bootstrapping. The χ^2 workflow used a 1+20-bit configuration with ring dimension 8192 and depth 2, with a 1+25-bit, depth 3 fallback in the threshold setting for error handling.

For Wilcoxon, the ranking primitive from [24] requires approximately $n \times n$ slots. Because CKKS exposes only $N/2$ usable slots, the ring dimension used here leaves insufficient precision headroom for stable downstream operations beyond $n = 64$, even though the ranking primitive itself can in principle support longer vectors. This implementation limit is one reason the Wilcoxon results are interpreted cautiously in Section 6.

5 Results

5.1 Ciphertext runtimes

Tables 2, 3, and 4 report core encrypted evaluation runtimes. These timings cover the homomorphic evaluation pipeline itself and exclude hospital-side ETL, ciphertext upload, governance approval latency, and threshold-release coordination. The relative ordering is consistent with the circuit structure: the rank-based Wilcoxon pipeline is dominated by comparison depth and bootstrapping, Pearson occupies the middle ground because of its masked client-aided finish, and χ^2 is the lightest routine.

5.2 Numerical accuracy

Table 5 summarizes the relative deviation between ciphertext and plaintext outputs. The encrypted χ^2 routine remained within 0.5% of the plaintext baseline across all tested cohort sizes. Pearson correlation stayed between 0% and 5.7% for $n \leq 256$, and the tuned higher-depth $n = 512$ run yielded 5%. Wilcoxon showed the largest deviation, with an approximately proportional inflation or attenuation of the z statistic.

Table 2 Pearson correlation ciphertext runtimes (ms).

Samples	4	8	16	32	64	128	256	512
1-party (FHE)	300.14	404.13	483.79	539.84	614.95	691.08	743.70	806.99
2-party (THE)	332.09	404.40	509.33	554.28	636.75	738.98	758.50	829.41
3-party (THE)	342.17	447.15	489.84	597.52	674.63	708.47	781.17	846.97

Table 3 Wilcoxon rank-sum ciphertext runtimes (ms).

Samples	4	8	16	32	64	128
1-party (FHE)	257.97	303.80	355.06	400.78	436.59	–
2-party (THE)	267.38	331.93	374.95	449.87	483.81	–
3-party (THE)	288.43	351.50	402.05	463.47	488.13	–

5.3 Approximate communication and public-key-material footprint

Table 6 summarizes the approximate external communication and public-key-material footprint for each statistic in the threshold-HE workflow. The figures are conservative engineering estimates derived from the representative CKKS contexts described in Section 4 and are intended to reflect external payloads at the architecture boundary, not internal ciphertext expansion during evaluation. Because the three routines use different parameterizations, the footprint is reported per statistic rather than under a single shared crypto context. For the estimates below, a fresh CKKS ciphertext is approximated as $2NL \times 8$ bytes, where N is the ring dimension and L is the active number of CRT primes in the relevant input context.

The dominant bandwidth cost is the one-time distribution of public evaluation material during setup, not the per-query upload from the participating centers. The upload and release sizes reported below therefore separate the routine architecture-boundary payloads from the public key material that is amortized across queries bound to the same key version.

5.4 Cross-platform runtime projection

Table 7 provides a cross-platform runtime projection anchored to the laptop environment used for the primary measurements and a server-class platform to give a preliminary sense of deployment-relevant performance. The laptop is a MacBook Pro with Apple M2 Pro / 12 cores (8P+4E), ARM NEON 128-bit SIMD, and 16 GB unified memory. The server platform is an AMD EPYC™ 8224P with 24 cores / 48 threads (Zen 4c, 2.55 GHz base / 3.0 GHz boost), AVX-512 SIMD, 64 MB L3 cache, and 6-channel DDR5 memory.

The EPYC values are projected from architectural comparison rather than direct measurements. The primary performance differentiator is AVX-512: the 512-bit datapath accelerates the Number Theoretic Transform (NTT) operations that dominate CKKS polynomial arithmetic by roughly 3–4× compared to the 128-bit ARM NEON path on the M2 Pro. The EPYC additionally provides 2× the physical core count for parallelizable sections of the OpenFHE pipeline (key generation, rotation-key evaluation). Combining the SIMD and threading advantages against the M2 Pro's higher per-core clock and IPC yields an estimated overall speedup factor of approximately 2.5–3× for these workloads. The reported ranges in Table 7 correspond to the 3-party threshold setting across the tested cohort sizes ($n = 4$ to $n = 512$ for Pearson and χ^2; $n = 4$ to $n = 64$ for Wilcoxon).

6 Discussion

These results show feasability for χ^2 and Pearson in the given setting. For this use case, χ^2 is the most numerically stable routine of the three,

Table 4 χ^2 ciphertext runtimes (ms).

Samples	4	8	16	32	64	128	256	512
1-party (FHE)	46.59	53.79	59.08	65.57	75.78	82.00	83.47	91.09
2-party (THE)	74.12	83.37	85.72	88.05	111.26	113.99	119.21	129.29
3-party (THE)	82.34	88.97	97.46	100.94	111.36	119.38	129.41	140.17

Table 5 Relative error (%) of ciphertext output versus plaintext baseline.

Samples	4	8	16	32	64	128	256	512
Pearson (%)	0.00	0.00	5.70	2.60	1.50	2.90	3.00	**5.00**
Wilcoxon z (%)	0.00	18.90	19.40	21.90	13.50	–	–	–
χ^2 (%)	0.00	0.13	0.18	0.29	0.18	0.17	0.23	0.46

Note: Pearson $n = 512$ reports the tuned higher-depth run. An initial depth-8 configuration produced a 60% outlier that is attributed to noise accumulation and is not used in the final summary.

and Pearson correlation is workable within the reported error range when parameters are tuned to the target cohort size. These two routines already cover common clinical questions about categorical association and linear dose-response behavior. They do not establish that arbitrary clinical statistics are now practical under threshold HE, but they do show that routine cross-site inference can sometimes be moved into an encrypted workflow without collapsing accuracy.

For a fixed consortium operating under a semi-honest model and explicit decryption governance, threshold CKKS is already useful for a narrow class of routine clinical statistics.

A further limitation is that the multi-center setting was emulated by partitioning a single 51-patient clinical cohort into site-specific subsets. Accordingly, the present evaluation does not empirically validate heterogeneous inter-hospital ETL, schema drift, cross-site latency, or live institutional governance workflows. These remain deployment questions for future work.

Wilcoxon is the clear weak point. The main issue is not the final threshold decryption step; it is the encrypted ranking and comparison pipeline. The observed 13.5–21.9% error in the reported z statistic means that near-threshold findings should be treated as exploratory, not definitive. This is especially important in small clinical cohorts, where significance margins are already fragile before encryption is introduced. For the present workflow, Wilcoxon is therefore better understood as a demonstration of current limits than as a ready-for-routine production statistic.

Governance is part of the technical result, not a separate administrative appendix. Once client-aided release enters the workflow, security depends on who may request a decryption, how requests are bound to a dataset and key version, how often similar decryptions are allowed, and whether the output can be audited later. In a hospital consortium, these controls are as relevant as ring dimension or multiplicative depth because they determine whether the privacy model survives contact with real operations. That is why the paper emphasizes workflow, architecture, governance, and the hospital-side integration boundary together, rather than presenting ciphertext evaluation in isolation.

Threshold HE is not the only option for collaborative analytics. Table 8 summarizes the main trade-offs against other privacy-enhancing technologies commonly discussed in biomedical data science [5–7]. The comparison is intentionally qualitative. The goal is not to crown one PET as universally best, but to clarify where threshold HE fits: a good option for fixed consortia that want strong protection of data in use with limited disclosure of final outputs, but that can tolerate substantial compute cost and formalize decryption governance.

Table 6 Approximate external communication and public-key-material footprint per statistic in the threshold-HE workflow.

Statistic	Representative CKKS context	Input ciphertexts / center	Center → computation service (per center)	One-time public eval. material	Computation service → threshold decryption	Threshold decryption → lead analyst	Logical passes after upload
Pearson correlation	upper-end tuned THE context; $N = 16384$, depth 8, 9 CRT primes (fresh ciphertext ≈ 2.25 MB)	2 $(\vec{x}, \vec{y})$	≈ 4.5 MB	≈ 100–200 MB	3 masked-result ciphertexts (≈ 6.75 MB)	3 masked scalars; analyst reconstructs r locally	2
Wilcoxon rank-sum	reported THE context; $N = 16384$, depth 10 with bootstrapping (fresh ciphertext ≈ 2.75 MB)	2 $(\vec{x}, \vec{g})$	≈ 5.5 MB	≈ 250–350 MB	2 masked-result ciphertexts (≈ 5.5 MB) + interactive bootstrap-share traffic	bootstrap shares + 2 masked scalars; analyst reconstructs z locally	4+
χ^2	conservative THE fallback context; $N = 8192$, depth 3, 4 CRT primes (fresh ciphertext ≈ 0.50 MB)	2 $(\vec{t}, \vec{\ell})$	≈ 1.0 MB	≈ 40–80 MB	8 masked-result ciphertexts (4 numerator/denominator pairs; ≈ 4.0 MB)	8 masked scalars; analyst reconstructs cell contributions and sums χ^2 locally	2

Note: Values are approximate external payloads, not measured wire traces. The "Logical passes" column counts architecture-level communication phases after input upload, not per-share-holder broadcast duplication. The Pearson row is intentionally conservative because smaller Pearson runs use lighter contexts. The χ^2 row uses two input vectors only $(\vec{t}, \vec{\ell})$, not three. In the χ^2 client-aided phase, the release consists of four masked numerator/denominator pairs, i.e. eight masked terms, rather than four cell counts.

Table 7 Cross-platform runtime projection (3-party threshold setting, ms). EPYC values are analytical projections based on architectural scaling; they are not measured benchmarks.

Platform	CPU / cores	RAM	Pearson (ms)	Wilcoxon (ms)	χ^2 (ms)
MacBook Pro	Apple M2 Pro / 12 cores (8P+4E)	16 GB	342–847	288–488	82–140
AMD EPYC 8224P (proj.)	Zen 4c / 24 cores (48 thr.)	$\geq$64 GB	$\approx$115–280	$\approx$95–165	$\approx$27–47
Projected speedup			$\approx$3.0×	$\approx$3.0×	$\approx$3.0×

Note: The Apple M2 Pro / 12 cores (8P+4E) row summarizes the range of 3-party runtimes from Tables 2–4. The EPYC projection assumes AVX-512 acceleration of NTT kernels ($\approx$3–4× over NEON) partially offset by the M2 Pro's higher single-core frequency. Actual server measurements may differ depending on compiler flags, OpenFHE threading configuration, and memory subsystem behavior under load.

The approximate footprint analysis in Section 5.3 confirms that per-center upload volumes remain modest (about 1.0–5.5 MB), while one-time public evaluation material dominates the setup exchange (about 40–350 MB depending on the statistic). The cross-platform runtime projection in Section 5.4 suggests that server-class hardware with AVX-512 can reduce ciphertext runtimes by roughly 3× relative to the laptop baseline, bringing even the heavier Pearson and Wilcoxon pipelines well under one second. The substantive conclusion remains unchanged: threshold HE is already useful for selected hospital-consortium statistics, but adoption still depends on key custody, ETL discipline, institutional coordination, and better support for rank-based inference.

7 Conclusion

This paper presents an applied workflow, architecture, and governance model for privacy-preserving multi-center clinical analytics using threshold CKKS and OpenFHE. The contribution is not a new cryptographic primitive. For a fixed hospital consortium, it is a concrete combination of encrypted statistical evaluation, a narrow hospital-side integration boundary, and controlled output

Table 8 Qualitative comparison of privacy-enhancing technologies for collaborative clinical analytics.

Technique	Main trust assumption	Primary bottleneck	Strength in clinical collaboration	Main limitation / caveat	Best-fit scenario
Threshold HE	Untrusted compute server; fewer than threshold key-share holders collude	Ciphertext computation, bootstrapping, key management	Strong protection of patient-level data in outsourced query execution; limited interaction during evaluation	Expensive arithmetic; onboarding and output governance are operationally demanding	Fixed consortium with predefined analytics and strict limits on plaintext release
Secure multi-party computation	Computing parties do not exceed the protocol's collusion bound	Multi-round interaction and network coordination	Exact collaborative statistics without exposing raw inputs to other sites	High interaction cost and orchestration burden across parties	Settings with stable multi-party coordination and strong need for exact arithmetic
Trusted execution environments	Hardware root of trust and correct attestation; side channels are acceptably mitigated	Enclave memory, attestation, hardware exposure	Near-plaintext performance and broad function support	Hardware trust, side-channel risk, and platform dependence	Workloads that prioritize performance and can accept hardware-based trust assumptions
Differential privacy	Noise budget is acceptable for the release goal	Privacy-utility trade-off and budget management	Strong formal protection for released summaries or models	Accuracy degradation can be severe in small cohorts or borderline inference	Public or semi-public release of aggregate outputs rather than exact joint computation
HE + MPC hybrid	Trust and collusion assumptions are split across both components	Combined engineering complexity	Can place different subroutines on the mechanism that fits them best	More difficult implementation, testing, and governance	Complex workflows where one PET alone is too rigid or too costly

release for a pediatric craniospinal irradiation use case.

Among the three implemented routines, χ^2 was the most numerically stable, Pearson correlation was workable within the reported error range, and Wilcoxon remained the fragile case because encrypted ranking still consumes too much precision and depth for comfortable routine use. The workflow therefore supports some clinically relevant cross-site statistics today, while also making clear where caution is required.

The main next steps are straightforward: more stable comparison and ranking methods, a broader statistical toolbox, stronger protections for malicious settings, and fuller deployment evaluation that includes communication, storage, and additional hardware platforms. Those steps would improve the practical maturity of the system, but the current results already show that patient-level data do not need to be centralized in plaintext for every clinically useful consortium analysis. The resulting contribution is therefore an applied workflow and consortium architecture for bounded clinical analytics.

Statements and Declarations

Funding

The authors received no financial support for the research, authorship, and publication of this article.

Competing interests

The authors declare that they have no competing interests.

Author contributions

G.C.T. drafted the manuscript. A.C.T. verified the medical framing and clinical interpretation against the original study. C.L. implemented the code and carried out the experiments. G.G. and M.T. supervised the work, refined the technical framing, and

contributed to manuscript revision. All authors contributed to study conception, approved the final manuscript, and agree to be accountable for the work.

Ethics and data handling

The patient cohort from [12] was used under the approvals governing the original clinical study. For the encrypted workflow studied here, plaintext patient data remained inside the originating institution's secure environment, and cross-site computation was performed on encrypted values only. The present manuscript focuses on the secure analytics workflow and does not introduce new patient recruitment, new intervention, or new biomedical claims beyond the original study. In view of the above and of the restrospective nature of this work, ethical approval was waived by the local Ethics Committee of The Oncology Institute Prof. Dr. Ion Chiricuţă.

Data availability

Patient-level clinical data are not publicly available because they contain sensitive health information and remain subject to ethical, institutional, and data-protection restrictions. Requests for access may be considered by the corresponding author and the originating institution, subject to applicable approvals and data-sharing requirements.

Code availability

The implementation used in this study is publicly available at https://github.com/certfhe/fhe_medical_insights.

References

[1] Narayanan A, Shmatikov V. Robust Deanonymization of Large Sparse Datasets. In: 2008 IEEE Symposium on Security and Privacy; 2008. p. 111–125.

[2] Mascalzoni D, Paradiso A, Hansson MG. Rare disease research: breaking the privacy barrier. Applied & Translational Genomics. 2014;3(2):23–29. https://doi.org/10.1016/j.atg.2014.04.003.

[3] Austin CP, Cutillo CM, Lau LPL, Jonker AH, Rath A, et al. Future of rare diseases research 2017–2027: an IRDiRC perspective. Clinical and Translational Science. 2018;11(1):21–27. https://doi.org/10.1111/cts.12500.

[4] Randall S, Brown A, Ferrante A, Boyd J, Robinson S. Implementing privacy preserving record linkage: insights from Australian use cases. International Journal of Medical Informatics. 2024;191:105582. https://doi.org/10.1016/j.ijmedinf.2024.105582.

[5] Ficek J, Wang W, Chen H, Dagne G, Daley E. Differential privacy in health research: a scoping review. Journal of the American Medical Informatics Association. 2021;28(10):2269–2276. https://doi.org/10.1093/jamia/ocab135.

[6] Cho H, Froelicher D, Dokmai N, Nandi A, Sadhuka S, Hong MM, et al. Privacy-Enhancing Technologies in Biomedical Data Science. Annual Review of Biomedical Data Science. 2024;7:317–343. https://doi.org/10.1146/annurev-biodatasci-120423-120107.

[7] Ballhausen H, Corradini S, Belka C, Bogdanov D, Boldrini L, Bono F, et al. Privacy-friendly evaluation of patient data with secure multiparty computation in a European pilot study. npj Digital Medicine. 2024;7:280. https://doi.org/10.1038/s41746-024-01293-4.

[8] Lee CH, Lim KH, Eswaran S. A comprehensive survey on secure healthcare data processing with homomorphic encryption: attacks and defenses. Discover Public Health. 2025;22:137. https://doi.org/10.1186/s12982-025-00505-w.

[9] Geva R, Gusev A, Polyakov Y, Liram L, Rosolio O, Alexandru A, et al. Collaborative privacy-preserving analysis of oncological data using multiparty homomorphic encryption. Proceedings of the National Academy of Sciences of the United States of America. 2023;120(33):e2304415120. https://doi.org/10.1073/pnas.2304415120.

[10] Lu Y, Tian Y, Zhou T, Zhu S, Li J. Multicenter Privacy-Preserving Cox Analysis Based

on Homomorphic Encryption. IEEE Journal of Biomedical and Health Informatics. 2021;25(9):3310–3320. https://doi.org/10.1109/JBHI.2021.3071270.

[11] Lee Y, Seo J, Nam Y, Chae J, Cheon JH. HEaaN-STAT: A Privacy-Preserving Statistical Analysis Toolkit for Large-Scale Numerical, Ordinal, and Categorical Data. IEEE Transactions on Dependable and Secure Computing. 2024;21(3):1224–1241. https://doi.org/10.1109/TDSC.2023.3275649.

[12] Turcas AC, Homorozeanu B, Gheara C, Balan C, Cosnarovici R, Diaconu O, et al. Dynamics and predictors of hematologic toxicity during cranio-spinal irradiation. Reports of Practical Oncology and Radiotherapy. 2024;29(3):362–372. https://doi.org/10.5603/rpor.101094.

[13] Cheon JH, Kim A, Kim M, Song Y. Homomorphic Encryption for Arithmetic of Approximate Numbers. Advances in Cryptology – ASIACRYPT 2017. 2017;p. 409–437. https://doi.org/10.1007/978-3-319-70694-8_15.

[14] Al Badawi A, Alexandru A, Bates J, Bergamaschi F, Cousins DB, Erabelli S, et al. OpenFHE: Open-Source Fully Homomorphic Encryption Library. Cryptology ePrint Archive. 2022;2022:915.

[15] Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. Journal of Biomedical Informatics. 2014;50:234–243. https://doi.org/10.1016/j.jbi.2014.04.003.

[16] Paddock S, Abedtash H, Zummo J, Thomas S. Proof-of-concept study: homomorphically encrypted data can support real-time learning in personalized cancer medicine. BMC Medical Informatics and Decision Making. 2019;19:255. https://doi.org/10.1186/s12911-019-0983-9.

[17] d'Aliberti OG, Clark MA. Preserving patient privacy during computation over shared electronic health record data. Journal of Medical Systems. 2022;46:85. https://doi.org/10.1007/s10916-022-01865-5.

[18] Lu Y, Zhou T, Tian Y, Zhu S, Li J. Web-Based Privacy-Preserving Multicenter Medical Data Analysis Tools Via Threshold Homomorphic Encryption: Design and Development Study. Journal of Medical Internet Research. 2020;22(12):e22555. https://doi.org/10.2196/22555.

[19] Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, Cuendet MA, Sousa JS, Cho H, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. Nature Communications. 2021;12:5910. https://doi.org/10.1038/s41467-021-25972-y.

[20] Kim E, Jeong J, Yoon H, Kim Y, Cho J, Cheon JH. How to Securely Collaborate on Data: Decentralized Threshold HE and Secure Key Update. IEEE Access. 2020;8:191319–191329. https://doi.org/10.1109/ACCESS.2020.3030970.

[21] Mouchet C, Bertrand E, Hubaux JP. An Efficient Threshold Access-Structure for RLWE-Based Multiparty Homomorphic Encryption. J Cryptol. 2023;36(10). https://doi.org/10.1007/s00145-023-09452-8.

[22] Kang DHE, Kim D, Song Y, et al. Harnessing the Potential of Shared Data in a Secure, Inclusive, and Resilient Manner via Multi-Key Homomorphic Encryption. Scientific Reports. 2024;14:13626. https://doi.org/10.1038/s41598-024-63393-1.

[23] Cheon JH, Kim D, Kim D. Efficient Homomorphic Comparison Methods with Optimal Complexity. In: Advances in Cryptology – ASIACRYPT 2020; 2020. p. 221–256.

[24] Mazzone F, Everts M, Hahn F, Peter A. Efficient Ranking, Order Statistics, and Sorting under CKKS. In: 34th USENIX Security Symposium (USENIX Security 25); 2025. p. 8541–8558. Available from: https://www.usenix.org/conference/usenixsecurity25/presentation/mazzone.

[25] Bost R, Popa RA, Tu S, Goldwasser S. Machine Learning Classification over Encrypted Data. In: 22nd Annual Network

and Distributed System Security Symposium (NDSS). Internet Society; 2015. p. 1–14.

[26] Lu Wj, Huang Z, Hong C, Ma Y, Qu H. PEGASUS: Bridging Polynomial and Non-Polynomial Evaluations in Homomorphic Encryption. In: 2021 IEEE Symposium on Security and Privacy (SP); 2021. p. 1057–1073.

[27] Hee SW, Willis A, Tudur Smith C, Day S, Miller F, et al. Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of ClinicalTrials.gov. Orphanet Journal of Rare Diseases. 2017;12:44. https://doi.org/10.1186/s13023-017-0597-1.

[28] Geroldinger M, Verbeeck J, Hooker AC, Thiel KE, Molenberghs G, et al. Statistical recommendations for count, binary, and ordinal data in rare disease cross-over trials. Orphanet Journal of Rare Diseases. 2023;18:391. https://doi.org/10.1186/s13023-023-02990-1.

[29] Xu W, Li X, Su Y, Wang B, Zhao W. Verifiable privacy-preserving cox regression from multi-key fully homomorphic encryption. Peer-to-Peer Networking and Applications. 2024;17:3182–3199. https://doi.org/10.1007/s12083-024-01740-9.