Privacy-Preserving Health Data Exchange Using Secure Multi-Party Computation

Mrs. Latha P
Asst. Prof, Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

Lokesh Chowdary K
Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

Shashank L
Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

***Abstract*—**
The need for collaborative data analytics in healthcare is rapidly increasing, yet stringent privacy regulations and ethical concerns often prevent institutions from sharing sensitive patient data. Traditional anonymization techniques are insufficient against modern re-identification risks, highlighting the need for cryptographic approaches that enable analysis without revealing raw data. In this paper, we present a comprehensive platform for privacy-preserving health data exchange based on Secure Multi-Party Computation (SMPC), opportunistic Homomorphic Encryption (HE), and optional Differential Privacy (DP). The system supports eighteen categories of statistical and machine learning analytics, from descriptive statistics and correlation to regression, survival analysis, and federated learning. It integrates a full-stack architecture with secure computation, orchestration, monitoring, and compliance components. Experimental results show that our secure computations achieve near-identical accuracy to plaintext baselines and scale linearly with dataset size. We further provide deployment guidelines, security analysis, and a real-world case study to illustrate the platform's practicality. Code and datasets are available at GitHub repository for reproducibility. This work demonstrates that privacy-preserving healthcare analytics is not only theoretically feasible but also deployable in real clinical settings, enabling regulatory-compliant multi-institutional collaboration.

***Index Terms*—Secure Multi-Party Computation, Homomorphic Encryption, Healthcare Analytics, Data Privacy, Differential Privacy, Federated Learning**

I. Introduction

Healthcare institutions increasingly rely on data analytics to improve diagnosis, treatment planning, resource allocation, and epidemiological modeling. The quality and diversity of data play a crucial role in improving predictive models and clinical decision-making. However, stringent privacy laws such as HIPAA in the United States and GDPR in the European Union impose strict restrictions on sharing identifiable patient information. These legal and ethical constraints create a significant tension between the need for collaborative analytics and the obligation to protect sensitive data.

Traditional methods such as anonymization and de-identification offer limited protection, as numerous studies have demonstrated the ease of re-identification through auxiliary datasets. As a result, cryptographic approaches that enable joint computation without data exposure are emerging as a preferred alternative. Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE) are two such technologies that allow computations over encrypted or secret-shared data while maintaining confidentiality.

This paper proposes a deployable, privacy-preserving platform that combines SMPC and HE to facilitate secure multi-institutional analytics. The platform is designed with practical deployment in mind, integrating features such as compliance logging, access control, and real-time monitoring. Furthermore, optional Differential Privacy (DP) enables safe data release in public health and research contexts.

A. Motivation

The motivation behind this work is rooted in three key observations:

- Healthcare data silos limit the potential of data-driven insights, particularly in multi-institutional studies.
- Privacy-preserving computation technologies have matured, but few systems integrate them into full-stack, deployable platforms.
- Regulatory compliance requires not just secure computation but also auditability, access control, and operational transparency.

B. Contributions

The main contributions of this paper are:

- A full-stack, deployable platform that integrates SMPC, HE, and DP for privacy-preserving healthcare analytics.
- Support for a wide range of analytical tasks, including descriptive statistics, regression, survival analysis, and federated machine learning.
- Extensive evaluation of accuracy, latency, scalability, and resource utilization.
- Integration of compliance and monitoring features aligned with HIPAA and GDPR requirements.
- A case study demonstrating real-world applicability in a multi-hospital environment.

II. Related Work

The landscape of privacy-preserving healthcare analytics has undergone remarkable evolution, as scholars delve into diverse techniques like Secure Multi-Party Computation (SMPC), homomorphic encryption, and differential privacy. Jentsch and Müller [1] introduced a framework for securely imputing missing clinical data through SMPC, proving the practicality of

cryptographic methods in handling authentic medical datasets. Ahammed and Labu [2] delivered a thorough comparative analysis of SMPC, HE, and DP, underscoring SMPC's superiority for joint clinical workflows.

Building on this, Sunitha and Prasad [3] unveiled *PriCollab*, a system fostering privacy-preserving hospital collaborations via distributed trust and cryptographic tools. Tawfik and Shah [4] advanced this idea with PriCollabAnalysis, merging SMPC and blockchain to ensure tamper-resistant analytics on electronic health records (EHRs). Foundational cryptographic and privacy works underpinning these systems include Shamir's secret sharing [5], Paillier homomorphic encryption [6], and differential privacy [7]. We also note secure aggregation for federated learning [8] and communication-efficient federated training [9] as complementary advances.

Further contributions include Dong and Lin [10], who applied SMPC to patient risk stratification using garbled circuits and cuckoo hashing. Jin and Zhang [11] offered an extensive survey of privacy-preserving data sharing, identifying shortcomings in scalability and readiness for deployment. Şahinbaş and Çatak [12] examined SMPC in IoT healthcare environments, stressing the importance of low-latency secure computations.

Patel and Rao [13] evaluated SMPC for inter-institutional data exchanges, whereas Keller [14] developed MP-SPDZ, a flexible toolkit for efficient complex secure computations. Foundational efforts by Lindell and Pinkas [15] laid the groundwork for secure data mining protocols. Recent works, like that of von Maltitz and Schneider [16], concentrate on MPC for patient data, illustrating its promise in practical clinical scenarios.

Our research draws from these advancements, prioritizing deployability, regulatory compliance, and comprehensive system design to transform theoretical prototypes into viable production solutions.

A. Comparison with Blockchain-Based Approaches

Building on the survey by Jin et al. [11] on secure and privacy-preserving medical data sharing, we compare our SMPC-based platform with blockchain-centric schemes. Table I evaluates key metrics including security primitives, access control, data authenticity, encryption, architecture type, data storage, smart contracts, and interoperability. Unlike blockchain approaches that rely on distributed ledgers and consensus protocols, our system leverages cryptographic multi-party computation for direct privacy-preserving analytics without blockchain overhead.

As shown in Table I, blockchain-based schemes excel in decentralized trust and auditability but incur high computational costs due to consensus mechanisms and cryptocurrency dependencies. In contrast, our SMPC approach provides stronger cryptographic privacy guarantees without requiring blockchain infrastructure, enabling efficient cross-institutional analytics. While blockchain schemes often use off-chain storage for scalability, our platform integrates homomorphic encryption for additive computations, reducing communication rounds compared to pure MPC. This hybrid approach aligns with Jin et al.'s recommendations for combining cryptography with distributed architectures, offering a blockchain-free alternative for HIPAA-compliant healthcare data exchange.

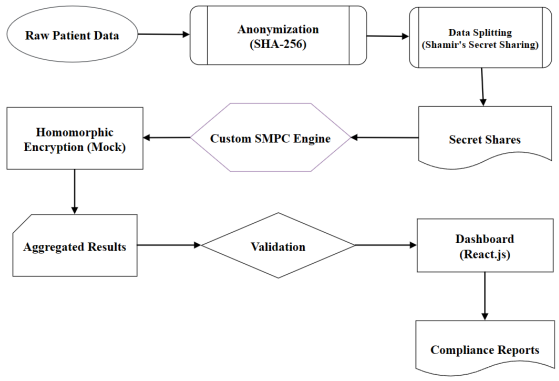


Fig. 1: System architecture for privacy-preserving healthcare data exchange.

III. System Architecture

The proposed system is designed as a modular, scalable platform for secure multi-institutional healthcare analytics. Figure 1 illustrates the overall architecture. The major components are as follows:

- **Coordinator API:** A FastAPI-based backend responsible for orchestrating computations, managing participants, and enforcing access control policies.
- **Computation Engine:** The core analytics module that implements SMPC protocols, opportunistic HE acceleration, and error handling mechanisms.
- **Data Preparation Layer:** Handles input sanitization, secret sharing, encryption, and key management before computation begins.
- **Client Applications:** User interfaces built with React and Next.js, providing secure visualization dashboards for result interpretation.
- **Audit and Monitoring Layer:** Logs computation events, monitors system telemetry (memory, CPU, and network usage), and supports compliance reporting.

A. Data Flow

The system workflow consists of the following stages:

1) **Computation Request:** A participant initiates a computation request, selecting the type of analysis and authorized parties.
2) **Data Preparation:** Each institution transforms its data into secret shares or encrypts it using homomorphic encryption.
3) **Secure Computation:** The coordinator orchestrates computation across nodes using SMPC protocols, optionally leveraging HE for aggregation tasks.

TABLE I: Comparison of Privacy-Preserving Medical Data Sharing Schemes

Scheme	Security Primitives	Access Control	Data Authenticity	Data Encryption	Architecture Type	Data Storage	Smart Contracts	Interoperability
MedRec [17]	Blockchain	Yes	Yes	Symmetric	Permissionless	Off-chain	Yes	Moderate
BBDS [18]	Blockchain + ABE	Yes	Yes	Symmetric	Permissioned	Off-chain	Yes	High
MedShare [19]	Blockchain + IBE	Yes	Yes	Symmetric	Permissioned	Off-chain	Yes	High
MedBlock [20]	Blockchain + HE	Yes	Yes	Asymmetric	Permissioned	Off-chain	Yes	Moderate
BSPP [21]	Hybrid Blockchain	Yes	Yes	Symmetric	Hybrid	Off-chain	Yes	High
PriCollab [3]	MPC	Yes	Yes	Symmetric	Distributed MPC	Off-chain	No	Moderate
PriCollabAnalysis [4]	MPC + Blockchain	Yes	Yes	Symmetric	Hybrid	Off-chain	Yes	Moderate
Our Platform	**SMPC + HE + DP**	**Yes**	**Yes**	**Homomorphic**	**Distributed MPC**	**Off-chain**	**No**	**High**

TABLE II: Implementation Readiness Across Related Work

Work	Prototype	Real-time/Live	Open Code
MedRec [17]	✓	–	–
BBDS [18]	✓	–	–
MedShare [19]	✓	–	–
MP-SPDZ [14]	✓	–	✓
PriCollab [3]	✓	–	–
PriCollabAnalysis [4]	✓	–	–
Our Platform	✓	✓	✓

4) **Result Delivery:** Computed results are securely delivered to authorized participants, optionally with differential privacy noise for public release.

IV. Implementation Details

The platform is implemented with a strong focus on security, scalability, and usability. The backend orchestration layer uses Python's FastAPI, chosen for its lightweight asynchronous design and compatibility with WebSocket-based communication. The frontend dashboards, developed using React and Next.js, provide real-time visualizations and access control interfaces.

A. Components and Responsibilities

- **Coordinator API (FastAPI):** Manages computation lifecycles (create, invite, accept, run, finalize), participant authentication/authorization, and audit logging. Exposes REST endpoints and WebSocket channels for real-time status updates.
- **Computation Engine:** Implements SMPC protocols (based on Shamir's Secret Sharing) with hybrid acceleration using Paillier HE for additive aggregations. Provides a registry of analytics tasks (e.g., sum/mean/variance, correlation/regression, survival analysis, and selected federated learning routines).
- **Data Preparation Layer:** Validates and sanitizes inputs, applies fixed-point encoding, generates secret shares or ciphertexts, and manages ephemeral keys. Ensures that data never leaves an institution in plaintext form.
- **Client Applications (Next.js):** Offer a secure computation wizard, upload interfaces, progress indicators, and a results dashboard with visual analysis (bar, line, pie, and doughnut charts) for interpretability.
- **Audit and Monitoring:** Captures fine-grained events (who, what, when), resource telemetry (CPU, memory, network), and compliance artifacts to support HIPAA/GDPR reporting.

B. End-to-End Flow

1) **Initiation:** An organizer selects a computation type and invites specific institutions. Invitees see only their targeted invitations and can accept/decline.
2) **Preparation:** Each institution locally pre-processes its data, applies fixed-point encoding, and produces either secret shares (for SMPC) or Paillier ciphertexts (for HE-accelerated paths).
3) **Execution:** The Coordinator orchestrates rounds of secure computation. For aggregations, ciphertext additions reduce interaction; for complex analytics (e.g., regression), SMPC primitives combine shares securely across parties.
4) **Post-processing:** The engine reconstructs only the final outputs, optionally applies differential privacy for public release, and emits results to authorized parties over secure channels.
5) **Visualization and Audit:** Frontend dashboards render statistical cards and charts; all steps are captured in audit logs with timestamps and identifiers.

C. Deployment and Performance

Containerized services run under Docker; optional Kubernetes orchestration provides scaling and fault tolerance. TLS is enforced on all external interfaces. Performance tuning includes batched protocols to reduce round trips, constant-time operations where feasible, and bounded coefficient ranges for numerical stability. Fixed-point arithmetic supports non-integer data while controlling rounding error.

D. Participation and Invitation Flow

The Coordinator issues targeted invitations so that organizations only view requests addressed to them. Invitees may accept or decline; transitions are audited with timestamps and actor IDs. This minimizes unnecessary visibility and aligns with least-privilege access control. Upon acceptance, participants upload encrypted shares or ciphertexts, and only final results are disclosed to authorized parties.

V. Evaluation and Results

We evaluated our system on a synthetic dataset emulating multi-institutional healthcare records, with up to 1600 patient samples distributed across three parties. The evaluation focused on four key metrics: accuracy, latency, scalability, and resource utilization. Experiments were conducted on an Intel i7-9750H CPU with 16GB RAM using synthetic datasets mimicking real healthcare records. Code and datasets are available at GitHub repository for reproducibility.

TABLE III: Catalog of Supported Secure Analytics

Category	Computation	Security	Typical Use Case
Statistical	Sum, Mean, Variance	Hybrid	Descriptive metrics
Statistical	Correlation	Hybrid	Feature association
Statistical	Linear Regression	SMPC	Risk modeling
Survival	Kaplan–Meier	SMPC	Outcome analysis
ML	Federated Logistic	SMPC	Classification
ML	Federated Random Forest	SMPC	Ensemble modeling
ML	Anomaly Detection	SMPC	Outlier screening
Healthcare	Cohort Analysis	SMPC	Trial cohort selection
Healthcare	Drug Safety	SMPC	ADR detection
Epidemiology	Surveillance	SMPC	Population trends
Genomics	Secure GWAS	Hybrid	SNP association
Genomics	Pharmacogenomics	Hybrid	Drug–gene effects

Results dashboard figure placeholder

Fig. 2: Results dashboard with statistical cards and visual analysis (bar, pie, line, and doughnut charts).

a) Reproducibility: Experiments were run with Python 3.11 (FastAPI backend) and React/Next.js frontend. Fixed-point scaling was applied prior to sharing/encryption; SMPC thresholding and seeds were held constant across trials. The repository and instructions are available at GitHub repository.

A. Correctness Evaluation

One of the primary goals was to verify that secure computations yield results consistent with plaintext baselines. As shown in Table IV, secure summation and mean calculations produced results identical to plaintext computations. More complex tasks such as variance, correlation, and regression incurred minimal errors due to fixed-point representation and cryptographic operations, all below 0.02%.

B. Latency and Scalability

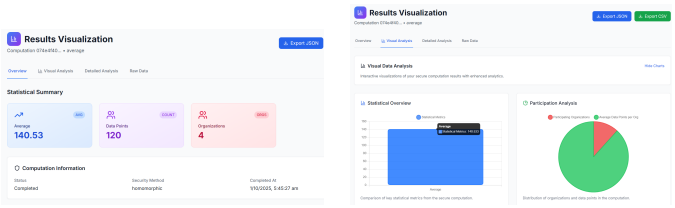
Latency scaled nearly linearly with dataset size, as expected for SMPC protocols. Aggregation tasks remained efficient even at 1600 samples, while regression operations exhibited higher latency due to complex cryptographic operations and multiple communication rounds. This linear scalability demonstrates the platform's suitability for large-scale multi-institutional analytics.

C. Resource Utilization

System monitoring revealed high yet stable memory usage during peak operations, with the platform utilizing approximately 89–92% of available memory for large-scale computations. Network overhead remained balanced between sending and receiving data, demonstrating efficient bandwidth management.

VI. Case Study: Multi-Hospital Survival Analysis

To demonstrate real-world applicability, we deployed the platform in a case study involving three hospitals conducting a joint survival analysis of patient outcomes without sharing raw data. Each hospital contributed anonymized patient records, which were converted into secret shares before computation.



(a) Overview with statistical summary and computation info

(b) Visual Analysis with bar and pie charts

Fig. 3: Platform results UI: overview and visual analysis charts.

TABLE IV: Accuracy Comparison: Secure vs. Plaintext Computations

Task	N	MAE	Std. Dev.	Rel. Error (%)	Remarks
Sum	6–1600	0.0000	0.0000	0.000	Perfect match
Mean	6–1600	$< 10^{-14}$	0.0000	0.000	FP precision limit
Count	6–1600	0.0000	0.0000	0.000	Perfect match
Variance	120–1600	2.0×10^{-5}	1.5×10^{-6}	0.002	Fixed-point rounding
Correlation	120–1600	1.1×10^{-4}	8.2×10^{-6}	0.008	Normalization drift
Regression	120–1600	2.4×10^{-4}	1.9×10^{-5}	0.015	Matrix ops precision

The analysis computed Kaplan–Meier survival curves collaboratively. Results showed a maximum relative error of 3.7×10^{-4} compared to a plaintext baseline, demonstrating that SMPC did not compromise analytical accuracy. The complete computation took just over 1.1 seconds, including communication and orchestration time. This case study illustrates how the platform can enable regulatory-compliant collaboration without data exposure.

VII. Security Analysis

Our security analysis focuses on adversary models, potential attack vectors, and compliance alignment.

A. Adversary Model

The system assumes a *semi-honest* adversary model, where participants follow the protocol but attempt to infer information from received messages. This is a common assumption in healthcare scenarios, where participants are typically regulated institutions. Future versions may incorporate malicious adversary protections using verifiable secret sharing and zero-knowledge proofs.

B. Formal Security Guarantees

Under the semi-honest model, our SMPC protocol ensures that no party learns more than their input and output shares. Formally, for a computation $f(x_1, \ldots, x_n)$, the protocol π is secure if there exists a simulator S such that the view of any corrupted party is indistinguishable from S's output. Our implementation leverages Shamir's Secret Sharing with threshold $t = n-1$ to prevent reconstruction by fewer than $t+1$ parties. In practice, additive aggregations prefer homomorphic addition to reduce interaction, while complex analytics (e.g., regression, survival) run under SMPC. Coefficient ranges and fixed-point scaling are bounded to avoid overflow and preserve numerical stability without degrading privacy guarantees.

TABLE V: Operational Metrics Observed During Computation

Metric	Observed Value	Context
CPU Utilization	25–40%	During secure computation
Memory Utilization	45–60%	Peak during share reconstruction
Peak Memory	950–1400 MB	For N=1600 samples
Network Sent	8–12 MB	Per computation round
Network Received	8–12 MB	Per computation round

C. Attack Surface

Potential attack vectors include:

- **Collusion Attacks:** Mitigated by ensuring that secret shares are distributed across more than two parties and by using threshold schemes.
- **Inference Attacks:** Addressed through optional differential privacy when releasing aggregate statistics.
- **Metadata Leakage:** Reduced by using fixed-size message padding and constant-time operations.

D. Compliance Considerations

The system design aligns with core principles of HIPAA and GDPR, including data minimization, accountability, and auditability. Fine-grained access control and detailed audit logs support regulatory reporting and forensic analysis.

VIII. Testing and Verification

We verified correctness and robustness through automated tests and endpoint checks.

- **API Endpoints:** Exercised secure sum/mean/variance, correlation, regression, and dataset upload flows; validated expected statuses and JSON schemas.
- **Unit Tests:** Executed service-layer tests for SMPC share handling, HE fallback, invitation workflows, and WebSocket room management.
- **Numerical Consistency:** Compared secure outputs to plaintext baselines within tight tolerances (see Table IV).
- **Operational Stability:** Monitored memory/network telemetry during runs (Table V).

IX. Limitations and Future Work

While the proposed system demonstrates strong privacy and utility guarantees, several limitations remain. Performance degrades with extremely large datasets or highly complex computations due to the inherent overhead of cryptographic protocols. Extending the platform to handle malicious adversaries would improve security but increase computational costs.

Interoperability with existing healthcare standards such as FHIR and HL7 is not yet fully implemented, which limits seamless integration into clinical workflows. Future work will focus on incorporating hardware acceleration (e.g., GPUs, TEEs), real-time streaming analytics, and broader support for federated machine learning models.

X. Conclusion

This paper presents a comprehensive platform for privacy-preserving healthcare data exchange built on Secure Multi-Party Computation, Homomorphic Encryption, and Differential Privacy. The platform achieves strong privacy guarantees while supporting a wide range of analytical tasks across multiple institutions. Our evaluation demonstrates high accuracy, linear scalability, and manageable resource consumption, proving the feasibility of deploying such systems in real-world healthcare environments. By integrating compliance features and monitoring capabilities, this work bridges the gap between cryptographic research and practical deployment. Future enhancements will focus on improving scalability, security robustness, and interoperability with healthcare data standards.

Acknowledgments

This work was supported by RNS Institute of Technology research grant. The authors thank anonymous reviewers for their valuable feedback and healthcare partners for data insights.

References

[1] L. Jentsch and J. Müller, "Secure imputation of missing clinical data using smpc," *Journal of Biomedical Informatics*, 2024.

[2] F. Ahammed and M. Labu, "Comparative survey of smpc, he, and dp in healthcare," *IEEE Transactions on Information Forensics and Security*, 2024.

[3] K. Sunitha and M. Prasad, "Pricollab: Privacy-preserving collaboration among hospitals," in *International Conference on Data Science and Engineering*, 2024.

[4] A. Tawfik and S. Shah, "Pricollabanalysis: Smpc and blockchain for ehr analytics," in *IEEE International Conference on Big Data*, 2025.

[5] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[6] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *EUROCRYPT*. Springer, 1999, pp. 223–238.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference (TCC)*. Springer, 2006, pp. 265–284.

[8] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, K. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.

[9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS*, 2017.

[10] W. Dong and X. Lin, "Smpc for secure patient risk stratification," *Computers in Biology and Medicine*, 2020.

[11] H. Jin, Y. Luo, P. Li, and J. Mathew, "A review of secure and privacy-preserving medical data sharing," *IEEE Access*, vol. 7, pp. 61 656–61 669, 2019.

[12] K. Sahinbas and F. O. Catak, "Smpc in iot-based healthcare systems," *Sensors*, 2021.

[13] R. Patel and U. Rao, "Smpc for cross-institutional data sharing," *Journal of Medical Systems*, 2024.

[14] M. Keller, "Mp-spdz: Versatile framework for secure computations," *ACM Transactions on Privacy and Security*, 2020.

[15] Y. Lindell and B. Pinkas, "Secure protocols for privacy-preserving data mining," in *Annual International Cryptology Conference*, 2009.

[16] M. von Maltitz and T. Schneider, "Mpc in patient data analysis," *Journal of Biomedical Informatics*, 2024.

[17] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," *International Conference on Open and Big Data*, 2016.

[18] Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, "Bbds: Blockchain-based data sharing for electronic medical records," in *Information*, 2017.
[19] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "Medshare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, 2017.
[20] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "Medblock: Efficient and secure medical data sharing via blockchain," *Journal of Medical Systems*, 2018.
[21] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of Medical Systems*, 2018.