Journal of Medical Systems (2022) 46:85
https://doi.org/10.1007/s10916-022-01865-5

IMPLEMENTATION SCIENCE & OPERATIONS MANAGEMENT

Preserving Patient Privacy During Computation over Shared Electronic Health Record Data

Olivia G. d'Aliberti[1] · Mark A. Clark[1]

Received: 24 November 2020 / Accepted: 16 September 2022 / Published online: 20 October 2022

Abstract

Patient Electronic Health Records (EHRs) contain valuable clinical data that is useful for medical research and public health inquires. However, patient privacy regulation and improper resource sharing risks limit access to EHR medical data for research and public health purposes. In this paper, we introduce an end-to-end security solution that addresses both concerns and facilitates the sharing of patient EHR data over an unsecured third-party server using a leveled homomorphic encryption (LHE) scheme. Time testing for aggregating queries and linear computations was carried out using an HPE ProLiant DL580 Gen 10 server with an Intel Xeon Platinum 8280 Processor.

Keywords Homomorphic Encryption · Electronic Health Records · Protected Health Information · Personally Identifiable Information · Fast Healthcare Interoperability Resources

Introduction

The widespread adoption of certified Electronic Health Records (EHRs) over the last decade has created siloed repositories of patient healthcare data [1]. Currently, individual health providers maintain isolated collections of digitized patient records with limited system interoperability and few sharing capabilities with external research teams. The ability to share EHR data across healthcare silos would be useful for medical research and public health analysis as it would create more comprehensive patient data sets [2] making possible health data analysis studies across multiple providers. More data, from multiple EHRs, would give reseaerchers the ability to better analyze condition-specific clinical outcomes and improve patient healthcare outcomes [3, 4]. However, concerns about improper resource sharing and maintaining patient confidentiality have limited the adoption of inter-hospital patient EHR data systems.

Improper EHR resource sharing occurs when Personally Identifiable Information (PII) or Protected Health Information (PHI) is shared without patient approval. This could take the form of a data breach in which an external party gains access to the data by attacking a third-party server, like the AMCA data breach [5]. But it could also originate from a less malicious, but equally dangerous, loss of security control when sensitive healthcare records are shared with multiple parties [6]. In either case, the leakage threatens institution reputation and jeopardizes patient privacy. It also violates the HIPAA Privacy Rule's PHI protection guarantee and could constitute a failure to meet the HIPAA Security Rule [7].

Maintaining patient PHI and PII privacy ensures that an individual patient cannot be connected back to their medical history. Confidentiality is important because leaking sensitive information can lead to patient stigma, embarrassment, and discrimination [8]. The current practice to ensure PHI remains protected for medical research is de-identification of patient data. However, even with anonymized EHR data, sensitive information might be learned in combination with external knowledge [9]; studies have shown that using prescription records [10], diagnosis code data [11], genomic data with allele frequency [12], improperly published medical data [12], or naïve suppression [13] it is possible for an

This article is part of the Topical Collection on *Implementation Science & Operations Management.*

✉ Olivia G. d'Aliberti
olivia.daliberti@leidos.com

Mark A. Clark
mark.a.clark@leidos.com

[1] Leidos Inc, Arlington, VA, United States

attacker to determine either patient uniqueness, or worse, patient re-identification. This risk is especially prevalent for smaller patient samples [14], such as rare disease patients or new medication releases, for whom inter-hospital EHR data sharing would be most beneficial.

In order to ensure controlled access to patient data and appropriate levels of confidentiality while encouraging inter-hospital resource sharing, we propose a customized but flexible leveled homomorphic encryption system (LHE). Homomorphic encryption (HE) schemes allow for additive and multiplicative operations to be performed over encrypted data. This means that once decrypted, the result is comparable to the same operations performed over plaintext. The benefit of this system is that it guarantees that information remains secure through data transfer and computation. Using this scheme, researchers would be able to securely perform aggregating and linear operations over encrypted medical data gathered from multiple sources while maintaining patient PHI and PII privacy. Moreover, by treating HE as a component within a larger data security system – with de-identification of patient data and access control procedures – this system ensures high levels of patient protection while allowing approved researchers access to relevant medical data.

Related Work

Ensuring PHI data privacy and health care system security while also permitting sharing of and computation over healthcare data is an ongoing research and implementation challenge. The use of an additive partially homomorphic encryption (PHE) scheme to protect EHR data during computation and storage in an untrusted third-party cloud server has been discussed in [4, 15–17]. All systems described are limited to data aggregation through addition and are unable to perform multiplicative operations over the encrypted data.

A more flexible, leveled homomorphic encryption scheme, which allows for limited multiplicative operations, is employed for aggregate EHR data encryption using HEANN [18, 19] and SEAL [20, 21]. In [19] and [21], researchers employ LHE schemes to average pre-generated differentially private histograms of diabetes patient medical records in third-party cloud servers. Computation performed directly over encrypted patient data for secure, wearable mobile health technology is discussed in [22] using HElib [23]. Similarly, an LHE scheme is used in a near real-time ECG-data monitoring system to securely compute the average heart rate of a patient in a cloud computing environment in [24] and to carry out linear regression model predictions in [25].

Our work builds upon the LHE approaches discussed above, using nGraph-HE [26], an HE-secure graph compiler, to encrypt data selected directly from a synthetic EHR system, to securely share across a third-party server, to efficiently generate data statistics, and to perform linear computations. We use a privacy structure for EHR data-sharing similar to those proposed in both [19, 21], focus on direct computation over encrypted data as in [22, 24], and would be able to generate predictions for a patient as in [25]. Using this system, we allow researchers the flexibility to generate their own computational commands, as in [15], while ensuring patient privacy and institutional oversight.

Methods

Synthea Synthetic Patients and HL7 FHIR Standard

We considered a data-sharing scenario using synthetic HL7 FHIR resources constructed using Synthea™ stored in a PostgreSQL relational database [27, 28]. Synthea is an open-sourced, synthetic patient generator that models complete medical history, which we selected in order to mimic realistic patient data and associated health records. The records are stored as modular Fast Healthcare Interoperability Resources (FHIR) JSON objects, which allows users access to atomic data elements. Meaning that, with Fhirbase, an open-source tool to store and manage FHIR data in PostgreSQL [29], both objects and their component parts are searchable. Using the FHIR framework and Fhirbase tool, patient data relevant for research and not PII can be quickly extracted for de-identification and encrypted sharing using SQL commands; see example commands in Appendix A1. We generated and stored two databases of 10,000 patients in this format as our synthetic EHR System for HE-encryption and computational experiments.

Data Analysis

We chose to focus on computation scenarios where data is aggregated or a linear regression computation is performed. Aggregation represents the most necessary and computationally simple statistic for health care analytics. Moreover, many of the barriers to information exchange across EHR systems are in cases where aggregation analysis is needed and analytics are being provided by third-party companies [30]. Regression analysis is a commonly used predictive technique in data mining – for healthcare data, it is widely used for predicting the disease or survivability of a patient [31]. However, current EHR networks do not often take advantage of the technique, because existing firewall-based setups do not allow for continual data flow, even for

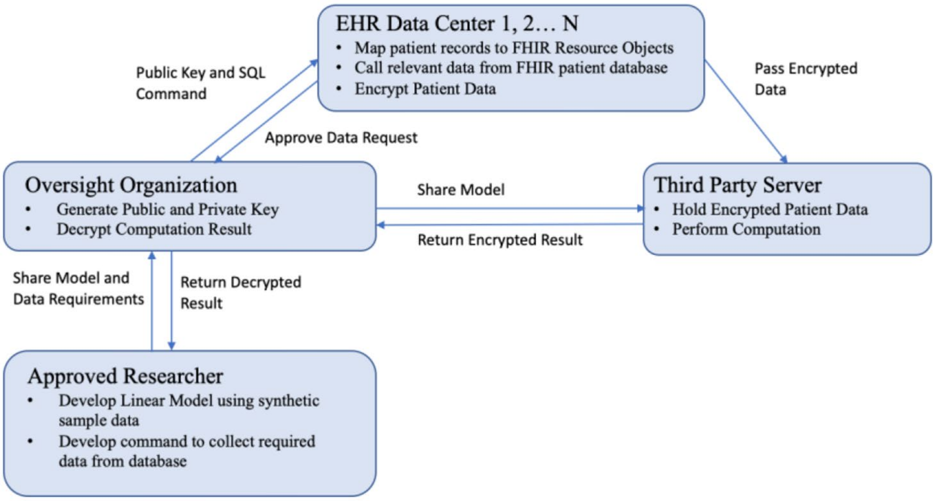

EHR Data Center 1, 2… N
• Map patient records to FHIR Resource Objects
• Call relevant data from FHIR patient database
• Encrypt Patient Data
Public Key and SQL Command
Approve Data Request
Pass Encrypted Data
Oversight Organization
• Generate Public and Private Key
• Decrypt Computation Result
Share Model
Return Encrypted Result
Third Party Server
• Hold Encrypted Patient Data
• Perform Computation
Share Model and Data Requirements
Return Decrypted Result
Approved Researcher
• Develop Linear Model using synthetic sample data
• Develop command to collect required data from database

Fig. 1 The proposed privacy preserving framework for data sharing

deidentified data, without human interference [32]. Our system would allow for both of these computation types to be set up and used across multiple EHRs in near-realtime.

Homomorphic Encryption

Public-key leveled homomorphic encryption (LHE) allows for depth-bound polynomial computation to be performed over encrypted data [33]. Using an LHE scheme, users are able to encrypt data with a public key, perform a limited number of additive and multiplicative operations over the encrypted data, return an encrypted response, and decrypt with a secret key to reveal the computational solution. The result is comparable to an answer obtained by performing the same calculate without encryption. For our research, we selected the CKKS scheme [34] as implemented by nGraph-HE [35]. We chose the CKKS scheme, because it allows for real-number calculations – meaning that test results could be directly stored and encrypted along with result code data. We selected nGraph-HE, which relies on Microsoft SEAL CKKS for underlying HE evaluation and nGraph for the graph compiler [36], because it has been optimized to perform LHE matrix computation over larger data sets.

System and Threat Model

Below is the model we propose for secure data transfer. There are three key entities including an: *Oversight Organization*, multiple *EHR Data Centers*, and the *Approved Researcher*. As in the diagram below, each entity would be responsible for the associated bullet-point tasks. We assume a semi-honest adversarial model – whereby attackers may attempt to gather information as available, but do not deviate from the protocol specifications.

Figure 1 above includes the following roles and associated responsibilities:

- *Approved Researcher*: A researcher or organization with approval to carry out computation on EHR data. Secondary analysis of existing data does not require Institutional Review Board (IRB) oversight if coded [37]; however, most institutions require internal approval to access data resources. Either the *Oversight Organization* or each individual *EHR Data Center* could be responsible for granting researcher approval. Once approved, the researcher is responsible for developing a standard data request. This reduces the workload on each *EHR Data Center* and ensures that all data is encrypted with the same underlying structure for computation. The researcher also develops the linear model for computation associated with their study. Some examples of requests that could be made include: an aggregation request, sample statistics generation, or a prediction given a multiple regression model.
- *Oversight Organization*: The oversight organization is a trusted party responsible for managing all data collection and computation request transfers for the *Approved Researcher* and the *EHR Data Centers*. It is their responsibility to verify that the data can be accessed and that neither the computation nor the requested data inappropriately reveal PII or PHI data. This is a necessary step in order to ensure data queried does not violate patient privacy by targeting individual or non-essential information in order to gain unapproved access to data. Once the *Oversight Organization* receives the data request format and computation from the *Approved Researcher*, they carry out the following steps: (1) approve researcher request; (2) create a public and private encryption key; (3) send the data request to each *EHR Data Center*

along with a public key for encryption; and (4) send the approved computation to the third party server. After computation is performed, the *Oversight Organization* decrypts with the private key and returns the unencrypted result to the approved researcher. We assume that the *Oversight Organization* maintains encryption protocol–meaning that there is no decryption prior to computation – and that it does not collude either with the *EHR Data Center* or the *Approved Researcher*.

- *EHR Data Centers*: The hospitals, physicians offices, or healthcare systems holding patient data. The *EHR Data Center* (1) responds to the data request; (2) encrypts the data using the public key; and (3) shares the information to the third-party server. *EHR Data Centers* can chose to not accept a query and can chose to only accept queries of certain types as determined in advance between the *Oversight Organization* and *EHR Data Center*.
- *Third Party Server*: Accepts the shared model, as determined by the researcher and passed to the oversight organization, and carries out computation over the aggregated LHE encrypted data passed from *EHR Data Centers*. The third-party server could be held locally by the oversight organization, but it could also be an unsecured cloud server. The benefit of using a cloud server is that the space for encrypted data and memory for computation is only needed while gathering encrypted data from the *EHR Data Centers* and executing the model. That data is never decrypted nor is it available within this server and thus is a limited security risk. Afterwards, the encrypted information could be deleted and the cloud server instance terminated.

Like in [18], this system relies on the *Oversight Organization* (*Trusted Third Party*) to manage key generation, oversee data collection from *EHR Data Centers*, and ensure the *Approved Researcher* is making appropriate requests. This layer of abstraction between hospital *EHR* and researchers is necessary in order to ensure the researcher cannot simply decrypt all stored data or that the computational queries neither gather nor reveal PHI/PII information. Many of the tasks the *Oversight Organization* carries out could be automated (key generation) or standardized (approved computational type). However, having an authoritative board stand between the *Approved Researcher* and the *EHR Data Centers* provides credibility and trust to the system. As such, we believe that an *Oversight Organization* controlling aggregation and computation is a necessary part of this data sharing model.

The benefit of this system is that nowhere, outside of the hospital *EHR Data Center*, is the raw patient data unencrypted. Patient data remains secure in transit and over computation.

That said, it is possible for an improperly managed/setup *Oversight Organization* to attempt to collude with the approved researcher in order to gain undue plaintext access to the data. The oversight organization could purposefully craft queries that have small enough data samples across a subset of hospitals in order to be reveal patient identity. In order to protect against this, we note that the system has been set up such that each *EHR Data Center* can approve/disapprove queries sent by the *Oversight Organization*, and moreover, in deployment suggest that appropriate query types be pre-determined by all parties before data is transferred.

Results

Prototype Implementation

We implemented the proposed solution as a client-server application in Python (version 3.5) and Tensorflow v1.14.0 for graph computation. For HE, we used nGraph-HE, the homomorphic encryption backend to Intel's nGraph compiler with a CKKS encryption scheme, as developed by Microsoft SEAL. Our module for EHR encryption allows for data extraction from a Synthea FHIR database using SQL commands, public key homomorphic encryption, and computation over encrypted data by a server. The client component enables users to encrypt across patient and event-level quantitative and one-hot encoded categorical variables using nGraph-HE. The server performs computation across over the available encrypted data, including aggregation and averaging and linear regression predictions.

Performance Evaluation

We tested performance of the proposed model on an HPE ProLiant DL580 Gen 10 server with an Intel Xeon Platinum 8280 Processor and an Ubuntu 16.04 operating system. Using a forked version of nGraph-HE, we simulated two different FHIR JSON databases and tested scenarios for encryption and computation. For each example computation, we discuss timing, storage, and error by size given different levels of security. All computations are measured by averaging over 10 trial runs at various computation size levels.

Aggregation.

In the first example case, we looked at SQL aggregation scenarios from the database. This work mimics the aggregating query setup in [14] and the histogram sharing in [3]. For example, we asked questions like:

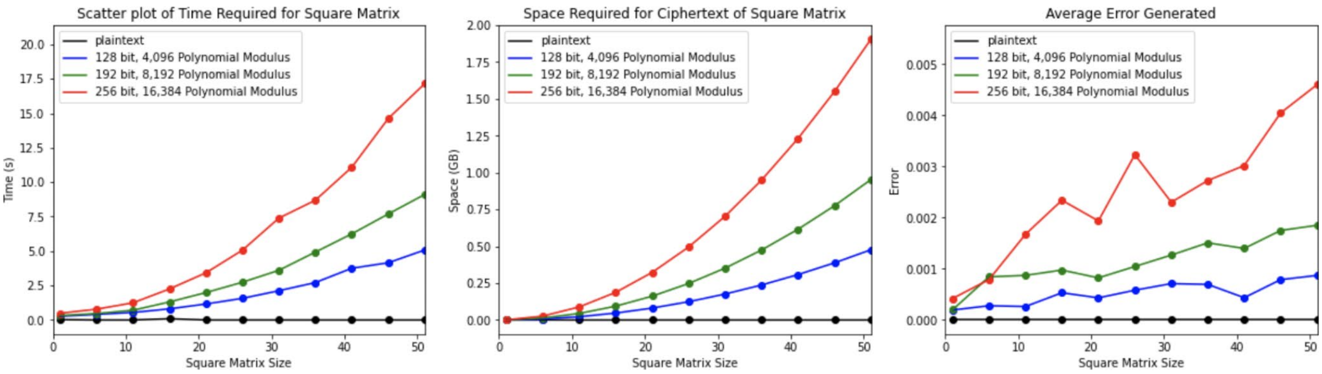


Fig. 2 All figures are averages of ten-run trials performed in square matrix size steps of five across plaintext, 128, 192, and 256-bt encryption levels using complex packing optimization

- 'How many patients prescribed drug X will have also received diagnosis Y in timeframe Z?'
- 'What is the average A for patients with condition B?'
- 'Given condition C, what percent of patients were given medications M_1, M_2, M_3?'

Each question is transformed into a SQL command to search the EHR FHIR JSONB database; see appendix A.1 for examples of SQL scripts that generate data capable of answering questions in the above format. Queries of this type meet the runtime, space, and error specifications shown in Fig. 2; primarily dependent upon input matrix size:

Once the queried data is returned, it is converted into a dataframe object. In order to process categorical data objects, like LOINC code names [38] and SNOMED Clinical Terms [39], variables are transformed into one-hot encoded vectors, whereby each unique value increases the size of the database by an additional vector. Quantitative data objects, like observational data and medicine dosage information, do not grow the dataspace in the same way, and are instead converted into tf.float32 compatible variables.

For each *EHR Data Center*, the resultant dataframe object, which could be a combination of one-hot encoded vectors and quantitative entries, is then flattened into an array, encrypted, and transferred along with shape to the server. The server waits until all data is collected from various pre-designated sources before carrying out computation. The data aggregation or averaging result, which would include data obtained from multiple *EHR Data Centers*, is then returned to the *Oversight Organization* for review and to the *Researcher* for analysis.

With increased matrix size, time and space constraints grow exponentially, while average error per input row remains constant, growing linearly over aggregation only with the increased number of row summations needed for larger square matrices. This is with the CKKS complex packing encoding optimization [40], which reduces memory and runtime constraints of HE. Without encryption,

Linear Regression Predictions.

In the second example, we looked at linear multiple regression prediction scenarios. Similar to cases discussed in [24], we focused on carrying out regression-style predictions for health care data. For example, we asked questions like:

- 'Given that patient A has observation B, what is test result should we expect for observation C?'
- 'What is the likelihood that patient will have disease D, given that they have test results T_1, T_2, T_3?'

We assume that researchers have pre-generated a linear model to use for predictions that can be shared with the server for prediction. An example use case for this scenario would be for a drug company carrying out post-market release testing on either a new treatment to (1) compare treatment effectiveness in pre-market release against newer results; (2) predict sample market-size across institutions; (3) determine additional use-cases based on information learned about patients in a post-market release.

Time required to run regression predictions on encrypted data depend upon number of variables included, the number of test samples, and the exponential power required by the regression equation. The more variables, the more test samples, and the higher the exponential power required for the regression equation, the longer it will take to return prediction(s). Moreover, in carrying out multiplicative operations over LHE encrypted data for linear and multiple regressions, it is important to be aware of ring size constraints to ensure your encrypted data remains within the cipherspace ring required for HE schemes. For a single variable linear regression equation of the format $y = \alpha x + \beta$ the following timing, space, and error constraints were found using our system with complex packing. Queries of this type meet the runtime, space, and error specifications shown in Fig. 3; and like aggregation, is also dependent upon input matrix size:

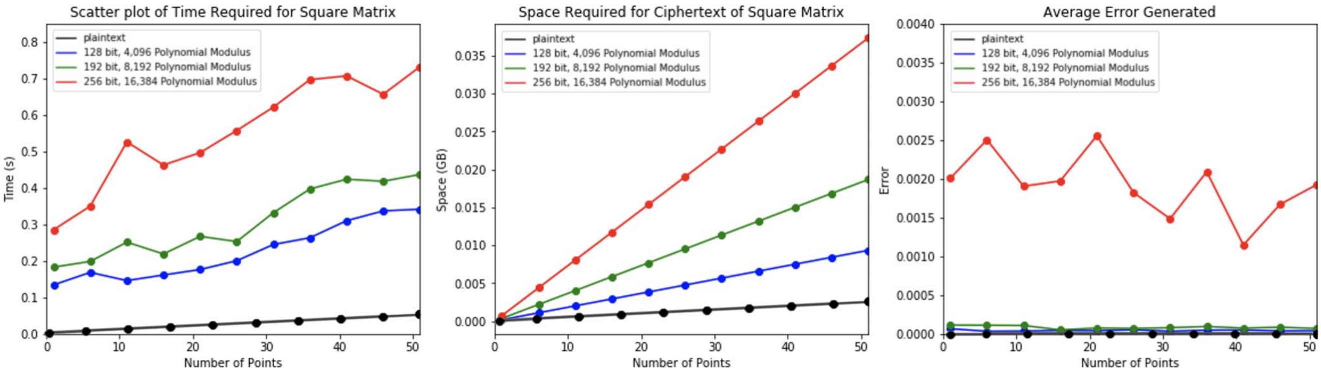


Fig. 3 All figures are averages of ten-run trials performed in point steps of 5 across plaintext, 128, 192, and 256-bit encryption levels using complex packing optimization

The space and time constraints grow linearly for small prediction samples, while absolute error fluctuates within the bounded, pre-set limits for each security level. Similar to the case above, categorical data is converted into one-hot encoded vectors, computation type is pre-generated, and data is aggregated at the server level prior to computation.

Conclusion

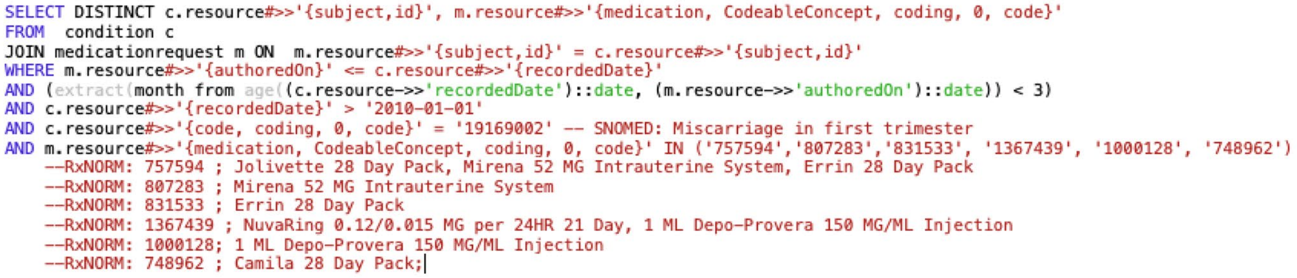
Patient EHR data is an important resource for research, and increased access would allow for a more accurate picture of public health, accelerate healthcare and medical research, and could bring treatments to patients sooner. However, due to silo-ed healthcare databases and an overarching concern for patient PHI and PII privacy, this resource has been largely untapped. We believe that a homomorphic encryption system could be a viable solution to this problem; providing data security during both transfer and computation. In particular, our work could allow researchers to collect resources across multiple institutions to generate aggregate statistics or make predictions for patients from multiple sources over an unsecured third-party server. A system like this, while slower than plaintext computation, makes available in near-real time a previously unattainable capability – secure data analysis while maintaining patient privacy and confidentiality.

Homomorphic encryption for EHR's has previously been seen as computationally restricted by both time and space constrains, and not practical due to EHR construction. However, recent research and implementation developments in EHR data storage and HE cryptographic schemes has made this type of construction more practical. In this paper, we have shown that given commercially available storage and power, it is feasible to carry out computation over EHR data for medical research. We believe that for small, disparate studies where patient privacy is of the utmost concern, a system like this could be invaluable. Using our system, LHE for EHR's is possible, and would be beneficial for computation across hospital systems for research purposes.

Appendix

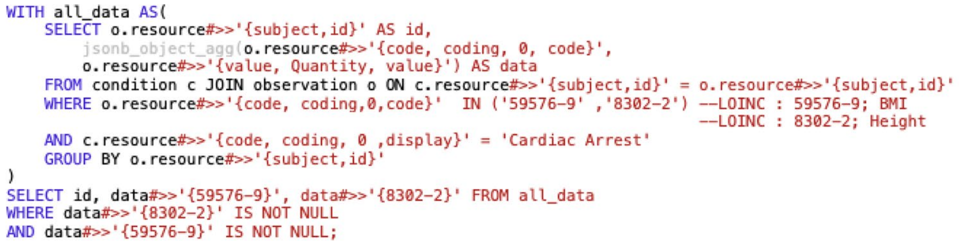
A.1 Sample PostgreSQL Data Requests

A.1.1 This JSONB SQL statement pulls records, from a FHIR health EHR database, of women who have suffered a miscarriage in the first trimester since 2010 and were prescribed a method of birth control 3 months prior to the miscarriage:



```
SELECT DISTINCT c.resource#>>'{subject,id}', m.resource#>>'{medication, CodeableConcept, coding, 0, code}'
FROM  condition c
JOIN medicationrequest m ON  m.resource#>>'{subject,id}' = c.resource#>>'{subject,id}'
WHERE m.resource#>>'{authoredOn}' <= c.resource#>>'{recordedDate}'
AND (extract(month from age((c.resource->>'recordedDate')::date, (m.resource->>'authoredOn')::date)) < 3)
AND c.resource#>>'{recordedDate}' > '2010-01-01'
AND c.resource#>>'{code, coding, 0, code}' = '19169002' -- SNOMED: Miscarriage in first trimester
AND m.resource#>>'{medication, CodeableConcept, coding, 0, code}' IN ('757594','807283','831533', '1367439', '1000128', '748962')
    --RxNORM: 757594 ; Jolivette 28 Day Pack, Mirena 52 MG Intrauterine System, Errin 28 Day Pack
    --RxNORM: 807283 ; Mirena 52 MG Intrauterine System
    --RxNORM: 831533 ; Errin 28 Day Pack
    --RxNORM: 1367439 ; NuvaRing 0.12/0.015 MG per 24HR 21 Day, 1 ML Depo-Provera 150 MG/ML Injection
    --RxNORM: 1000128; 1 ML Depo-Provera 150 MG/ML Injection
    --RxNORM: 748962 ; Camila 28 Day Pack;
```

A.1.2 This JSONB SQL statement pulls Body Mass Index and Height statistics, from a FHIR health EHR database, of individuals who have experienced a cardiac arrest episode:



```
WITH all_data AS(
    SELECT o.resource#>>'{subject,id}' AS id,
        jsonb_object_agg(o.resource#>>'{code, coding, 0, code}',
        o.resource#>>'{value, Quantity, value}') AS data
    FROM condition c JOIN observation o ON c.resource#>>'{subject,id}' = o.resource#>>'{subject,id}'
    WHERE o.resource#>>'{code, coding,0,code}'  IN ('59576-9' ,'8302-2') --LOINC : 59576-9; BMI
                                                                       --LOINC : 8302-2; Height
    AND c.resource#>>'{code, coding, 0 ,display}' = 'Cardiac Arrest'
    GROUP BY o.resource#>>'{subject,id}'
)
SELECT id, data#>>'{59576-9}', data#>>'{8302-2}' FROM all_data
WHERE data#>>'{8302-2}' IS NOT NULL
AND data#>>'{59576-9}' IS NOT NULL;
```

Funding This study was funded by Intel Corporation.

Compliance with Ethical Standards

Conflict of Interest Authors Mark Clark and Olivia d'Aliberti received an internal research and development (IR&D) grant from Intel Corporation.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Hartskamp, Michael Van, et al. "Artificial Intelligence in Clinical Health Care Applications: Viewpoint." Interactive Journal of Medical Research, vol. 8, no. 2, May 2019, doi:https://doi.org/10.2196/12100.
2. Garrett, Daniel. "Tapping into the value of health data through secondary use: as electronic health records (EHRs) proliferate across the nation, an important new opportunity awaits healthcare organizations that can find meaningful commercial uses for the data contained in their EHR systems." Healthcare Financial Management, vol. 64, no. 2, Feb. 2010, pp. 76.
3. Emam, Khaled El, et al. "A Secure Distributed Logistic Regression Protocol for the Detection of Rare Adverse Drug Events." Journal of the American Medical Informatics Association, vol. 20, no. 3, July 2012, pp. 453–461., doi:https://doi.org/10.1136/amiajnl-2011-000735.
4. Yadav, Pranjul, et al. "Mining Electronic Health Records (EHRs): A Survey"
5. AMCA Data Breach Impacts 12 Million Quest Diagnostics Patients." HIPAA Journal, 4 June 2019, https://www.hipaajournal.com/amca-data-breach-impacts-12-million-quest-diagnostics-patients/.
6. Abomhara, Mohamed, and Geir M. Køien. "Towards an Access Control Model for Collaborative Healthcare Systems." Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies, 2016, doi:https://doi.org/10.5220/0005659102130222.

7. The Health Insurance Portability and Accountability Act (HIPAA), (45 C.F.R. § 160, 164(a,e), 1996).
8. Nass, Sharyl J, et al. *Beyond the Hipaa Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, D.C: National Academies Press, 2009.
9. Li, Fengjun et al. "New Privacy Threats in Healthcare Informatics: When Medical Records Join the Web." 2010.
10. Emam, Khaled El, et al. "Evaluating the Risk of Re-Identification of Patients from Hospital Prescription Records." The Canadian Journal of Hospital Pharmacy, vol. 62, no. 4, 2009, doi:https://doi.org/10.4212/cjhp.v62i4.812.
11. Loukides, Grigorios, et al. "The Disclosure of Diagnosis Codes Can Breach Research Participants Privacy." Journal of the American Medical Informatics Association, vol. 17, no. 3, 2010, pp. 322–327., doi:https://doi.org/10.1136/jamia.2009.002725.
12. Thenen, Nora Von, et al. "Re-Identification of Individuals in Genomic Data-Sharing Beacons via Allele Inference." Bioinformatics, vol. 35, no. 3, 2018, pp. 365–371., doi:https://doi.org/10.1093/bioinformatics/bty643.
13. Vaidya, Jaideep, et al. "Identifying Inference Attacks Against Healthcare Data Repositories." *AMIA Joint Summits on Translational Science Proceedings.* 2013, pp. 262–66.
14. Courbier, Sandra, et al. "Share and Protect Our Health Data: an Evidence Based Approach to Rare Disease Patients' Perspectives on Data Sharing and Data Protection - Quantitative Survey and Recommendations." Orphanet Journal of Rare Diseases, vol. 14, no. 1, Dec. 2019, doi:https://doi.org/10.1186/s13023-019-1123-4.
15. Raisaro, Jean Louis, et al. *Feasibility of Homomorphic Encryption for Sharing I2B2 Aggregate-Level Data in the Cloud*. American Medical Informatics Association, 2017.
16. Ikuomola, Aderonke J. et al. "Securing Patient Privacy in E-Health Cloud Using Homomorphic Encryption and Access Control." International Journal of Computer Networks and Communications Security (IJCNCS) vol 2, January 2014, pp. 15–21.
17. Wang, Qi, et al. "Privacy Preserving Computations over Healthcare Data." 2019 *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2019, doi:https://doi.org/10.1109/ithings/greencom/cpscom/smartdata.2019.00123.
18. HEANN, https://github.com/kimandrik/HEAAN
19. "Privacy-Preserving Biomedical Data Dissemination via a Hybrid Approach." *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, vol. 2018, 2018, pp. 1176–85.
20. SEAL, https://github.com/Microsoft/SEAL
21. Chou, Edward, et al. A Fully Private Pipeline for Deep Learning on Electronic Health Records. Nov. 2018.
22. Preuveneers, Davy, and Wouter Joosen. "Privacy-Enabled Remote Health Monitoring Applications for Resource Constrained Wearable Devices." *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, vol. 04-08-, ACM, 2016, pp. 119–24, doi:https://doi.org/10.1145/2851613.2851683.
23. Shai Halevi and Victor Shoupn, https://github.com/shaih/HElib
24. Kocabas, Ovunc, et al. "Assessment of Cloud-Based Health Monitoring Using Homomorphic Encryption." 2013 *IEEE 31st International Conference on Computer* Design *(ICCD)*, IEEE, 2013, pp. 443–46, doi:https://doi.org/10.1109/ICCD.2013.6657078.
25. Bos, Joppe W., et al. "Private Predictive Analysis on Encrypted Medical Data." *Journal of Biomedical Informatics*, vol. 50, Elsevier Inc, Aug. 2014, pp. 234–43, doi:https://doi.org/10.1016/j.jbi.2014.04.003.
26. nGraph-HE, https://github.com/IntelAI/he-transformer
27. Walonoski, Jason, et al. "Synthea: An Approach, Method, and Software Mechanism for Generating Synthetic Patients and the Synthetic Electronic Health Care Record." Journal of the American Medical Informatics Association, vol. 25, no. 3, Oxford University Press, Mar. 2018, pp. 230–38, doi:https://doi.org/10.1093/jamia/ocx079.
28. Bender, Duane, and Kamran Sartipi. "HL7 FHIR: An Agile and RESTful Approach to Healthcare Information Exchange." *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, doi:https://doi.org/10.1109/cbms.2013.6627810.
29. Fhirbase, Health Samurai, https://github.com/fhirbase/fhirbase
30. Rahimzadeh, Vasiliki. "A Policy and Practice Review of Consumer Protections and Their Application to Hospital-Sourced Data Aggregation and Analytics by Third-Party Companies." *Front Big Data*, 2021, February. doi: 0.3389/fdata.2020.603044
31. Tomar, Diveya, et al. "A survey on Data Mining approaches for Healthcare." *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, 2013, pp. 241–266. doi: https://doi.org/10.14257/ijbsbt.2013.5.5.25
32. Shortreed, Susan M., et al. "Challenges and Opportunities for Using Big Health Care Data to Advance Medical Science and Public Health." American Journal of Epidemiology, vol. 188, no. 5, March 2019. doi: https://doi.org/10.1093/aje/kwy292
33. Gentry, Craig. "Fully Homomorphic Encryption Using Ideal Lattices." *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ACM, 2009, pp. 169–78, doi:https://doi.org/10.1145/1536414.1536440.
34. Cheon, Jung Hee, and Yong Soo Song. *Homomorphic Encryption Method of a Plurality of Messages Supporting Approximate Arithmetic of Complex Numbers*. 7 Feb. 2018.
35. Boemer, Fabian, et al. *nGraph-HE: A Graph Compiler for Deep Learning on Homomorphically Encrypted Data*. Oct. 2018.
36. Cyphers, Scott, et al. "Intel nGraph: An Intermediate Representation, Compiler, and Executor for Deep Learning." *arXiv.org*, Cornell University Library, arXiv.org, Jan. 2018, http://search.proquest.com/docview/2071286873/.
37. Coded Private Information or Specimens Use in Research, Guidance, Office for Human Research Protections (2008)
38. McDonald, Clem et al. "Introduction." LOINC Users' Guide (2017)
39. International Health Terminology Standards Development Organization. SNOMED CT® Editorial Guide, January 2020.
40. Boemer, Fabian, et al. "nGraph-HE2: A High-Throughput Framework for Neural Network Inference on Encrypted Data." *arXiv.org*, Cornell University Library, arXiv.org, Aug. 2019, http://search.proquest.com/docview/2272613116/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.