一般口演
# 一般口演11
# セキュリティとプライバシー保護2
2017年11月21日(火) 16:00 ～ 17:30 H会場 (10F 会議室1008)

[2-H-3-OP11-3] 分散医療情報分析基盤の秘密計算適用可能性の検証

木村 映善[1], 濱田 浩気[2], 諸橋 玄武[2], 千田 浩司[2], 岡本 和也[3], 真鍋 史郎[4], 武田 理宏[4], 三原 直樹[5], 黒田 知宏[3], 松村 志[4]
（1.愛媛大学大学院医学系研究科博士課程医学専攻 社会・健康領域医療情報学講座, 2.NTTセキュアプラットフォーム研究所, 3.京都大学医学部附属病院医療情報企画部, 4.大阪大学大学院医学系研究科医療情報学, 5.国立がん研究センター中央病院）

【背景】
i2b2は、各施設が公開可能なデータを各施設で保有・管理するアーキテクチャである。必要に応じて各所にクエリに集計結果を統合する仕組みが標準的に備えられている。さらにセキュリティを高めるために i2b2ノード群を仲介し、監査を実現する SHRINEも提案されている。しかし、 SHRINEは統計開示制限を行う前のデータを各ノードから Aggregatorに渡すために、 Aggregatorへの信頼が必要であること、自組織以外の組織群の結託によるデータの暴露に対して脆弱である。本研究は Aggregatorが不要な完全分散モデルにもとづく秘密計算を i2b2に実装し、実証実験を実施する。
【方法】
三大学に VPNで相互接続された i2b2ノードを設置し、合計3000人の患者、各患者に性別、年齢、 HBA1c、CR、 eGFR、尿蛋白、 CKD重症度、施設番号の8項目からなる24000件のダミーの observation factを投入した。筆者らが2015年に開発した SDC付き統計量の算出可能な秘密計算エンジンを i2b2各ノードに組み込んだ。
【結果】
最大値の計算では全体処理44.27秒、うちデータ読込17.81秒、相関係数・検定の計算では全体処理128.81秒、うちデータデータ読込16.73秒となった。 i2b2を介さず直接データを読み込んだ場合は0.1秒未満であり、データソースとしての i2b2のオーバーヘッドの大きさが確認された。
【考察】
i2b2の読出が遅いものの全体処理時間は実用的水準に達している。 i2b2のテーブル設計は多様なデータを保持できるようなカラム指向設計となっており、複数条件を指定したクエリに対して最大限の効率性を発揮するような設計になっていない可能性がある。 i2b2のクエリ最適化と実データの投入時の本邦の標準医療情報マスタとの整合性処理、性能評価等詳細な解析を進める予定である。

分散医療情報分析基盤の秘密計算適用可能性の検証

*木村映善[*1]、濱田浩気[*2]、諸橋玄武[*2]、千田浩司[*2]、岡本和也[*3]、
真鍋史郎[*4]、武田理宏[*4]、三原直樹[*5]、黒田知宏[*3]、松村泰志[*4]
*1 愛媛大学大学院医学系研究科博士課程医学専攻社会・健康領域医療情報学講座、
*2NTT セキュアプラットフォーム研究所、*3 京都大学医学部附属病院医療情報企画部、
*4 大阪大学大学院医学系研究科医療情報学、*5 国立がん研究センター中央病院

Feasibility study of applying secure multiparty computation on distributed medical information analysis system

*Eizen Kimura[*1], Kouki Hamada[*2], Gembu Morohashi[*2], Koji Chida[*2], Kazuya Okamoto[*3],
Shiro Manabe[*4], Toshihiro Takeda[*4], Naoki Mihara[*5], Tomohiro Kuroda[*3], Yasushi Matsumura[*4]
*1 Dept. Medical Informatics of Medical School of Ehime Univ., *2 NTT Secure Platform Laboratories,
*3 Kyoto University Hospital, Division of Medical IT & Administration Planning,
*4 Medical Informatics, Osaka University Graduate School of Medicine, *5 National Cancer Center Hospital

Abstract: i2b2 is the distributed architecture that the nodes at healthcare settings manage their health data with the mechanism of aggregating result from the nodes in the single query. SHRINE has been developed to enforce the audit and protection of privacy. However, in SHRINE, the nodes have to pass the data to Aggregator before processing statistical disclosure control. So, SHRINE is vulnerable to the exposure of data due to a collusion of groups of nodes other than the node. In this study, we implemented the secure computation based on a fully distributed model without an Aggregator and integrated i2b2. The 24000 dummy observation facts for 3000 patients consisting of items: Sex, Age, HbA1c, CR, eGFR, urine protein, CKD severity, medical institution number were loaded into the each i2b2 node. We incorporated the secure computation engine with SDC feature developed in 2015. For the calculation of the maximum value and the correlation coefficient of fact data, the whole process time were 44.27 and 128.81 seconds. The time taken to load the data were 17.81 seconds and 16.73 seconds. The overall processing time has reached a practical level. The table design of i2b2 is a column–oriented design that can hold various data, and there is a possibility that it may not be designed to demonstrate the maximum efficiency for queries specifying multiple conditions. We plan to carry out detailed studies; the optimization the performance of the query and database of i2b2, mapping the fact data with the Japanese standardized terminologies.

Keywords: secure computation, i2b2, big data, privacy, statistical disclosure control

1. はじめに

我が国は、データ駆動型医学(DDM : Data-Driven Medicine)[1]の研究活動を立ち上げ、Precision Medicine [2]の実現をめざす過渡期にある。DDM の実現には、横断的な分析が可能な状態の医療情報が多施設から入手できるようになっていることが前提となる。すなわち多施設間において、(1)臨床研究に用いる疾患・臨床情報・患者基本情報に関する医療情報モデルの構造化、(2)各施設が利用している医療情報モデル間のマッピング、あるいは標準化された医療情報モデルの共通採用、(3)医療情報の安全な共有と解析の方法論、といった諸技術を確立し、各拠点のデータを統合して運用できることが求められている。本研究では、(1)(2)の課題解決と密接に連携しつつ、主に(3)の医療情報の安全な共有と解析に関する技術の確立をめざす。

我が国では、多施設の臨床研究における研究者へのデータ共有・提供について二つの方向性が考えられる。一つは従来の臨床研究で取り組まれていた方式であり、各医療機関からデータを収集して臨床研究リポジトリに中央集権的に蓄積し、リポジトリから分析に必要なデータを提供する方法である。もう一つは、各医療機関のデータサイトを安全な医療ネットワークで接続し、研究に必要な医療情報に限定してオンデマンドで提供する方法である。

今回の概念実証は、後者の複数参加機関がデータを持ちよる形態を想定する。すなわち、次に述べる要素技術が必要である。(1)標準化された医療情報を保有する、データベース機能を有するリポジトリ群、(2)各リポジトリが保有するデータを必要以上に開示することなく、かつ特定個人を識別特定させずにリポジトリ群横断的に各々の個人に属するデータを集計あるいは統計処理を実現する技術。(3)集計されたデータを安全に開示できるような統計的開示制限あるいは匿名化処理技術、である。データを安全に集計処理するにあたって、(1)から(3)のアーキテクチャは密接に連携し、一体のシステムとして機能することが求められる。本研究は上記の要件を満たすシステムを模索すべく、プロトタイプの開発と評価を行う。

2 関連研究

2.1 標準医療情報を保有するリポジトリ

前項で述べたように、多施設連携のために標準化された医療情報をデータベース上に有するリポジトリを用意する。静的な医療情報の配置方法として、我が国では SS-MIX 標準化ストレージがあるが、現時点で外部からの動的なデータ検索・取り出し要求に対して標準化されたインタフェースは策定されていない。そこで医療情報の保存と検索について標準的インタフェースを提供しているリポジトリシステムの採用を検討した。

臨床的表現型を記録した医療データベースとゲノム情報とを統合したトランスレーショナルリサーチを実現す

るために、Informatics for Integrating Biology and the Bedside (i2b2)[3]プロジェクトが立ち上げられた。i2b2 はファクト情報と、そのファクト情報を修飾するデータ群のための標準化されたデータベーススキーマを Data Repository Cell (DRC)において定義している。但し、DRC に格納される情報が依存するオントロジーについて利用者が合意する必要がある。なお、ここで言及しているオントロジーとは、一般的なオントロジーではなく、i2b2 プロジェクトが、みずからの用語集の管理機構及びその管理機構によって管理されるターミノロジーの枠組み[4]をオントロジーと称しているものである。また、各組織が保有するデータの医療情報モデルから、DRC の スキーマに準拠するようにデータモデルの変換を行う必要がある。

2.2 分散クエリ環境

i2b2 は単体での稼働を想定して設計しており、複数施設の i2b2 ノードを統合して検索するフェデレーション技術は i2b2 本体では提供されてない。そのため、i2b2 の外部プロジェクト(The Shared Health Research Information Network (SHRINE))[5]が開発されている。SHRINE は i2b2 の各データベースにクエリを投入し、各ノードがクエリの結果を Aggregator に返し、Aggregator が統計的開示制御を施した最終的な結果をユーザに返す Trusted Third-party(TTP)信頼モデルを採用したシステムである。この方式の欠陥は、Aggregator が結果の一部分を知り得ること、自分以外のノードが結託した場合はデータが漏洩する可能性があることである。現在の情報漏洩のインシデントの多くは、データにアクセスできる者によるデータ持ち出しによる。故に、Aggregator のように、セキュリティ上の Single Point of Failure になるような TTP を介在しない、完全な分散モデルが望まれる。そこで、従来の秘密計算システムに加えて、クエリ要求に応えて各ノードにおいて i2b2 からデータを抽出して秘密計算に供するモジュールを開発する。

2.3 セキュリティ技術

秘密計算のアルゴリズムの選択にあたり、本概念実証実験では以下の条件を考慮した。(1)長期間にわたって医療情報を安全に保管できること、(2)データ量の増大に対して計算時間が有用な範囲に収まるような、スケーラビリティのある手法であること、(3)多数の参加者があるため、参加者の数だけデータ侵害に関するリスクが増大する。そのため、内部者による不適切な行為(データ侵害、漏洩)に対する耐性を有すること。

プライバシー保護のアプローチは大別して k-匿名化に代表される匿名化[6]・再構築法[7]と秘密計算に大別される。前者は後者の秘密計算と比較して計算コストが小さく、分析者にデータを渡した後のプライバシー侵害や情報漏洩に対する耐性が高い。しかし、データを非可逆な方法で加工することによるデータ精度の低下がある。また、個別の医療機関のデータを非特定識別状態で結合した上で匿名化処理するのではなく、個別の機関から匿名化された状態で収集する場合は、希少例は個別に抑制されて全体像から大きく外れた形で集計される可能性がある。

一方、秘密計算[8]は暗号化や秘密分散によってデータを秘匿化した上で、元のデータに戻さずに統計量を得ることができる性質を持ち、かつ匿名化のようなデータの変性をともなわないため、希少症例や機微性の問題からデータ提供が望ましくないデータに関して統計分析する用途に期待されている。秘匿化状態のデータを用いた計算を可能にする方法には、準同型暗号を用いる方法[9]と秘密分散を用いる方法[10]が知られている。

先述の判断基準を踏まえて、秘密分散方式にもとづいた手法を選択する。秘密分散方式は、準同型暗号より比較的計算量が抑えられる。また、Shamir の(K,N)閾値秘密分散[11]にもとづいた方式では、参加者 N 者のうち K-1 者以下のデータ漏洩、あるいは結託した開示に対する耐性を有する。そしてこの性質は計算量的安全性ではなく情報的安全性にもとづくため、暗号化解読に関する技術の進歩に対する耐性をも有する。

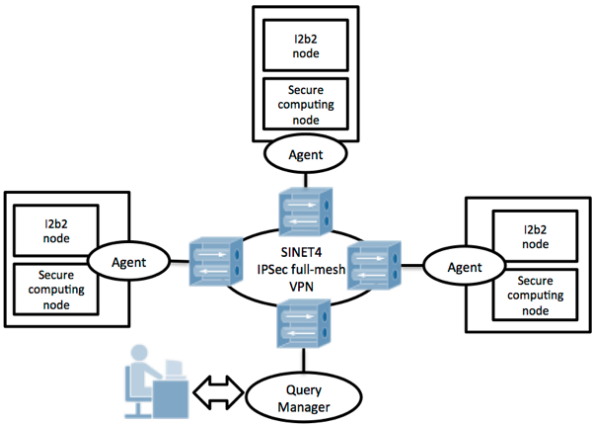

図 1 秘密計算ノードと併置した i2b2 ノード

3. 方法

3.1 秘密計算環境

各サイト（三大学）間を IPSec のフルメッシュ VPN で相互に接続する。各サイトに Damgård らの方式[12]を改良した秘密計算処理を実施する秘密計算ノード[13, 14]と i2b2 ノードを設置する（図 1）。データ分析者は Query Manager(QM)にデータの抽出条件、統計処理を要求する。以下、参加大学全体での性別毎の Cr の平均算出という単純な処理で本システムの動きを例示する（図 2）。ユーザは QM を通して、(1)各ノードの「クエリ・集約機能」に対してクエリを要求する。「クエリ・集約機能」は i2b2 へ性別、Cr に関してデータ抽出を要求し結果を受け取る。複数の結果が返されるので、性別、件数、Cr の総和に集約し、秘密計算ノードに取り込む。各ノードでの「クエリ・集約機能」の処理が完了したことを確認したのち、「秘密計算機能」に対して性別枚に Cr の平均値を算出する指示を出し、結果としての性別枚の Cr の平均値を得る。

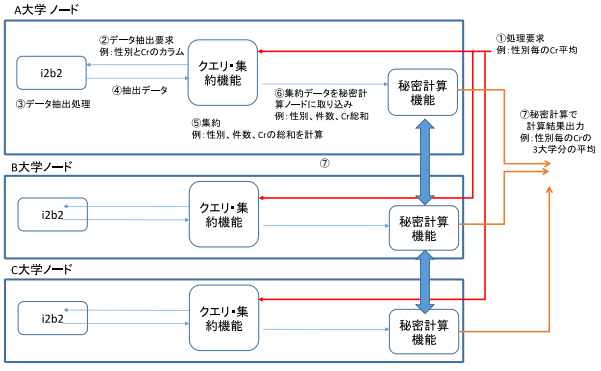

図 2 クエリ集約と秘密計算の処理フロー

3.2 統計的開示制限

データの開示制限はデータの利用者の制限と情報量の管理を組み合わせて行われる[15]。後者のデータ情報量の開示は、n 以下の個体数であれば公開を拒否することで、希少例の存在の類推を防ぐ Cell Suppression や、k 人以下の個人を区別できなくする k-匿名化などを通して、特定個人を識別可能な状態にする手がかりをなくす処理であり、統計的開示制限(Statistical Disclosure Control: SDC)[16]とよばれる。

SDC の閾値は、どの n-1 施設でも(各施設から返された度数の総和)≧t を満たすように設定した。総和に対して、m 施設でセル値の k%以上を占めている数量表のセルを隠蔽する占有ルールとして m=1,k=80%と設定した。すなわち, 全ての総和のセルにおいて、どの S(n-1 施設)と X(m 施設,X⊂S でも、(X の総和)/(S の総和) ≦ k/100 を満たすように制御される。

3.3 検証シナリオ

検証シナリオとして、HbA1、Cr(血清クレアチニン)、推定糸球体濾過量(eGFR)、尿蛋白, 慢性腎臓病の重症度の相関を確認する統計的検定を設定する。2012/04/01 から 2014/03/31 までの 3 年間に各大学医学部附属病院で検体検査が行われた 20 歳以上の患者を対象とした。さらに、全ての検査項目が 30 日以内に取得されていることを条件とし、年齢、性別、Hb1Ac、尿蛋白、Cr を電子カルテより抽出し, CSV ファイルに保存した。これらの CSV ファイルを各大学の i2b2 ノードに取り込む。基本的統計量に関する基本性能を検証するために, 患者を高齢者(65 歳以上)、非高齢者(20 歳以上 65 歳未満)、男性、女性で分類し, 男性、女性, 高齢者男性、高齢者女性, 非高齢者男性, 非高齢者女性の 6 種類のグループを作成し, 以下の統計解析を実施した。eGFR は 6 区間(15 未満, [15,30], [30,45], [45,60], [60,90], 90 以上)、HbA1c は 5 区間(6.2 未満, [6.2,6.9], [6.9,7.4], [7.4,8.4], 8.4 以上)に層別した。基本的統計量に関する基本性能を検証するために, eGFR の区間を横軸, HbA1c の区間を縦軸にしたマトリクスのセル毎に, Cr の統計値(総和、度数, 平均, 分散, 最小、最大, 中央値)を計算した。HbA1c の区間別に、Urinary Protein, CKD 重症度, eGFR に対して ANOVA, Kruskal-Wallis 解析を実施した。HbA1c の連続値と eGFR について Pearson correlation coefficient, および尿蛋白について Spearman rank correlation coefficient を算出した。

処理性能のプロファイリングを実施するために、3 種類の処理を設定した。

(処理 1)i2b2 を使用せずに、CSVファイルから読み取って秘密計算処理を実施する[14]。

(処理 2)i2b2 ノードにデータを読み込み、各条件に該当する observation fact を読み込み、その全てに含まれる患者の属性値を使用して処理する。

(処理 3) i2b2 ノードにデータを読み込み、各条件に該当する observation fact を読み込むが、前までの全ての条件を満たす患者IDに絞って読み込んでから処理する。

4. 結果

指定条件で抽出した結果、対象となった患者は愛媛大 619 名, 京都大 817 名, 大阪大 1517 名であった。 患者ごとに性別, 年齢, HbA1c, Cr 値, eGFR, 尿たんぱく, CKD 重症度, 施設番号の 8 項目からなるデータを i2b2 向けに展開し、i2b2 に延べ 24000 件の observation fact として投入した。

各試行の結果を表 1 に掲載する。各施設ノードに保有する患者数の差異から処理時間が異なるので、各施設の 2 回ずつの試行結果を提示している。「i2b2 なし」は、従来の i2b2 を介さない、秘密計算処理システムだけでの実証結果である。「i2b2-1」、「i2b2-2」は i2b2 ノードにデータを保有した状態で、全ての患者を対象とするか条件を満たす患者のみを対象としたのかの違いである。統計量の算出は、SDC なしと SDC ありについて個別に計算しており、その処理時間の総和を「合計時間」として計上している。「データ読み込み時間」は、その「合計時間」のうちに占める、各試行のデータ読み込み時間の総和であり、内訳として i2b2 と xml パースを提示している。「i2b2 なし」については i2b2 を使用していないため、データ読み込み時間の内訳はなく CSV データの読み込みの所要時間のみである。

i2b2-2 の SDC 付き統計量の計算(試行 2)では, 44.272 秒(3 拠点の最大値)の全体処理時間に対し, データ読み込みに最大で 17.806 秒を要した。3 拠点の最大値を採用した理由は、各ノードへの処理を同時に要請し、並列稼働して個別に計算が終了するために、最後に終了したノードの処理時間が全体の処理終了時間と同義であるからである。

i2b2-2 の相関係数・検定の計算(試行 2)では, 128.809 秒(3 拠点の最大値)の全体処理時間に対し, データ読み込みに最大で 16.730 秒を要した(うち 16.095 秒が i2b2 への問合せに要した時間)。いずれの場合も i2b2 なしのものではデータ読み込みに合計でも 0.1 秒に満たず, 秘密計算の処理時間に比べると無視できる程度と言えるが, i2b2 を使う場合はデータ読み込みの時間が全体の 1 割から 4 割くらいと必ずしも無視できない程度となった。

統計的開示制限は各ノードで適切に実施され、eGFR、HbA1c の外れ値における出現頻度が少ない検査数について統計的開示制限が機能し、適切に開示抑制がなされたことが確認された。また、秘密計算によって導出された統計量は、平文での統計処理の結果と一致していることが確認された[14]。

5. 考察

5.1 パフォーマンス

i2b2 の読み出し部分が遅いものの、全体の処理時間は十分実用的なレベルに達していると考える。i2b2 からの読み出しが遅いのは、i2b2 のデータベースはカラム志向であるが、カラム志向データベースシステムではなく、従来の RDBMS 上で実現しているため、処理が最適化されていない可能性がある。i2b2 の設計、データベースエンジンの特性についても検討して i2b2 の最適化を図ることが望ましいとおもわれる。

5.2 SS-MIX 標準化ストレージとの関係

本邦では医療情報は医療情報標準化推進協議会(HELICS)にて定められた医療情報標準化指針に掲載されている標準マスタ群に準拠するようになっており、それらの標準マスタ群に準拠したデータを標準化されたファイル配置方法で格納する方式として SS-MIX 標準化ストレージ（以下 SS-MIX）が普及している。本実験では、標準化マスタを利用した標準化メッセージ群が格納されている SS-MIX から i2b2 へのデータを取り込むという仮定で行う。すなわち、SS-MIX からデータを DRC に取り込むプログラムを開発することで、(i)DRC へのデータモデルの変換、(ii)組織横断的な i2b2 利用におけるオントロジーの合意が解決されるものとする。本実験は i2b2 と秘密計算の連携検証が主な目的であるため、今回は SS-MIX か

らのデータ抽出ではなく、参加医療機関の電子カルテから i2b2 に取り込める形式の XML データを生成し、i2b2 に取り込むものという仮定で進めた。今後は、実際に SS-MIX から i2b2 ノードデータへ取り込み、秘密計算を実行するまでの全プロセスを通した検証を検討している。

5.3 次世代医療基盤法における認定事業者との関係

本概念実証は、複数の組織によるデータ持ち寄りを想定して検証したが、データを中央に集める方式でも利用機会があると考える。次世代医療基盤法で規定されている認定事業者は、中央集権的にデータを収集する事業を執り行うものである。しかし、一つの認定事業者が国内に存在する医療情報について全ての種類、かつ全てのレコードを悉皆的に収集することは困難である。巨視的な視点に立つならば、全ての医療情報を一つの認定事業者で保有することは困難であるため、研究者の研究内容に応じて複数の認定事業者間で各々が保有している医療情報を組み合わせて、研究者が要求するデータセットにまとめ上げることも考えられる。認定事業者間で安全に医療情報を共有し、必要なデータ・統計量を抽出するのにも、この秘密計算が応用できることが期待される。

6. 結語

本研究では i2b2 という標準医療情報リポジトリ間における秘密計算の可能性を提示した。本研究では n 者が参加し、n-1 までの結託耐性を有する秘密計算のプロトコルを採用し、統計的開示制限も実装した。データ管理者からのデータ侵害, データ利用者による機微情報の推定を抑制し, かつ実際の臨床研究で使われている各種統計処理を実用的な計算速度で実行できることを示した.

i2b2 自体は特定の情報モデル、オントロジーに依存しない、列指向のデータベースとして構築されており、本邦の医療情報も標準マスタに則ったデータであり、データ較正が適切に行われているとするならば、特段の支障なく運用可能であることが示唆された。

7. 謝辞・COI

本研究は、「匿名化を適用した医療健康情報分析の有用性評価」の共同研究契約にもとづいて、NTT セキュアプラットフォーム研究所からの研究資金の提供下に NTT セキュアプラットフォーム研究所、愛媛大学、大阪大学、京都大学の四者間の共同実験下に実施されました。本実験にご協力頂いた関係者にこの場をお借りしてお礼を申し上げます。本研究の研究デザイン・研究活動は大学の研究者によって実施されており、研究スポンサーは患者データに関してなんら関与していません。共著者は自らが所属

表 1 各データ集約と秘密計算の処理時間

愛媛(患者数619) 各演算に要した時間(秒)

	SDCなし							SDC有り							データ読み込み時間の内訳			
	count	sum	mean	var	max	median	min	count(sdc)	sum(sdc)	mean(sdc)	var(sdc)	max(sdc)	median(sdc)	min(sdc)	合計時間	データ読み込(i2b2)		(xmlパース)
i2b2なし(試行1)	0.617	0.943	6.429	9.445	1.093	3.800	1.112	0.375	0.757	0.386	0.381	0.380	0.389	0.384	26.489	0.046		
i2b2-1(試行1)	9.083	3.835	8.734	11.438	3.013	6.279	3.064	1.870	2.199	1.859	1.847	1.990	1.807	1.896	58.914	20.008	19.737	0.198
i2b2-2(試行1)	1.776	2.663	8.042	10.946	2.614	6.027	2.400	1.655	2.184	1.533	1.758	1.582	1.679	1.636	46.497	13.584	13.312	0.133
i2b2なし(試行2)	0.612	0.947	6.424	9.444	1.098	4.314	1.103	0.386	0.747	0.384	0.387	0.378	0.384	0.384	26.992	0.047		
i2b2-1(試行2)	2.290	2.843	8.376	11.335	3.046	6.041	2.953	1.714	2.348	1.890	1.838	1.905	1.841	1.759	50.179	10.739	10.505	0.170
i2b2-2(試行2)	1.736	2.201	7.603	10.686	2.398	5.753	2.677	1.403	1.997	1.499	1.547	1.565	1.551	1.510	44.127	10.348	10.101	0.119

	pearson	anova	spearman	ruskal-wallis														
i2b2なし(試行1)	11.020	16.842	42.700	39.187											109.749	0.053		
i2b2-1(試行1)	13.249	23.393	44.467	46.214											127.323	9.990	9.729	0.185
i2b2-2(試行1)	13.256	23.269	47.227	45.639											129.391	9.581	9.314	0.159
i2b2なし(試行2)	11.025	16.855	43.627	38.625											110.131	0.051		
i2b2-1(試行2)	13.300	23.004	45.328	45.305											126.938	8.385	8.148	0.164
i2b2-2(試行2)	13.305	22.661	46.609	46.186											128.760	8.760	8.483	0.163

京都(患者数817) 各演算に要した時間(秒)

	SDCなし							SDC有り							データ読み込み時間の内訳			
	count	sum	mean	var	max	median	min	count(sdc)	sum(sdc)	mean(sdc)	var(sdc)	max(sdc)	median(sdc)	min(sdc)	合計時間	データ読み込(i2b2)		(xmlパース)
i2b2なし(試行1)	0.630	0.956	6.442	9.451	1.108	3.812	1.120	0.388	0.772	0.393	0.387	0.390	0.393	0.398	26.641	0.044		
i2b2-1(試行1)	9.099	3.845	8.742	11.446	3.021	6.285	3.070	1.881	2.200	1.870	1.862	1.994	1.818	1.909	59.042	28.946	28.566	0.295
i2b2-2(試行1)	1.791	2.676	8.052	10.956	2.623	6.032	2.405	1.670	2.194	1.548	1.764	1.588	1.691	1.643	46.633	18.026	17.606	0.218
i2b2なし(試行2)	0.628	0.953	6.439	9.458	1.113	4.327	1.109	0.397	0.762	0.395	0.395	0.391	0.395	0.392	27.154	0.056		
i2b2-1(試行2)	2.306	2.856	8.390	11.347	3.060	6.053	2.965	1.725	2.362	1.899	1.849	1.916	1.847	1.773	50.346	15.012	14.633	0.292
i2b2-2(試行2)	1.751	2.204	7.618	10.690	2.409	5.764	2.692	1.417	2.010	1.512	1.562	1.572	1.562	1.510	44.272	13.727	13.344	0.206

	pearson	anova	spearman	ruskal-wallis														
i2b2なし(試行1)	11.033	16.859	42.710	39.196											109.798	0.073		
i2b2-1(試行1)	13.265	23.404	44.475	46.221											127.365	13.872	13.500	0.269
i2b2-2(試行1)	13.266	23.275	47.237	45.657											129.434	12.462	12.028	0.266
i2b2なし(試行2)	11.038	16.860	43.644	38.638											110.180	0.060		
i2b2-1(試行2)	13.316	23.020	45.340	45.312											126.988	12.526	12.141	0.277
i2b2-2(試行2)	13.319	22.671	46.622	46.196											128.809	12.240	11.811	0.258

大阪(患者数1517) 各演算に要した時間(秒)

	SDCなし							SDC有り							データ読み込み時間の内訳			
	count	sum	mean	var	max	median	min	count(sdc)	sum(sdc)	mean(sdc)	var(sdc)	max(sdc)	median(sdc)	min(sdc)	合計時間	データ読み込(i2b2)		(xmlパース)
i2b2なし(試行1)	0.625	0.953	6.440	9.448	1.105	3.809	1.115	0.387	0.770	0.390	0.384	0.388	0.390	0.396	26.600	0.099		
i2b2-1(試行1)	9.095	3.843	8.739	11.440	3.021	6.282	3.067	1.878	2.197	1.867	1.859	1.991	1.815	1.905	58.998	24.623	24.115	0.420
i2b2-2(試行1)	1.784	2.673	8.049	10.953	2.620	6.029	2.402	1.663	2.188	1.546	1.760	1.585	1.688	1.640	46.581	17.806	17.237	0.318
i2b2なし(試行2)	0.622	0.950	6.436	9.456	1.110	4.320	1.106	0.391	0.759	0.393	0.392	0.389	0.392	0.389	27.106	0.100		
i2b2-1(試行2)	2.304	2.854	8.387	11.344	3.050	6.051	2.962	1.724	2.360	1.895	1.846	1.913	1.841	1.771	50.302	23.579	23.040	0.451
i2b2-2(試行2)	1.746	2.198	7.616	10.687	2.409	5.757	2.690	1.413	2.004	1.509	1.560	1.569	1.559	1.508	44.225	17.001	16.449	0.309

	pearson	anova	spearman	ruskal-wallis														
i2b2なし(試行1)	11.032	16.855	42.705	39.193											109.785	0.109		
i2b2-1(試行1)	13.261	23.401	44.468	46.215											127.345	17.970	17.438	0.405
i2b2-2(試行1)	13.257	23.270	47.232	45.652											129.411	16.726	16.106	0.392
i2b2なし(試行2)	11.038	16.853	43.637	38.636											110.164	0.103		
i2b2-1(試行2)	13.314	23.018	45.330	45.307											126.970	17.422	16.892	0.403
i2b2-2(試行2)	13.319	22.668	46.623	46.189											128.798	16.730	16.095	0.411

する医療機関のデータのみにアクセスを制限しています。

参 考 文 献

[1] Shah NH, Tenenbaum JD. FOCUS on translational bioinformatics: The coming age of data-driven medicine: translational bioinformatics' next frontier. Journal of the American Medical Informatics Association: JAMIA. 2012;19(e1):e2.

[2] Collins FS, Varmus H. A new initiative on precision medicine. New England Journal of Medicine. 2015;372(9):793-5.

[3] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association. 2010;17(2):124-30.

[4] Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. Journal of the American Medical Informatics Association. 2016;23(5):909-15.

[5] Weber GM, Murphy SN, McMurry AJ, MacFadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. Journal of the American Medical Informatics Association. 2009;16(5):624-30.

[6] Sweeney L. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002;10(05):557-70.

[7] Agrawal R, Srikant R, editors. Privacy-preserving data mining. ACM Sigmod Record; 2000: ACM.

[8] Yao AC, editor. Protocols for secure computations. Foundations of Computer Science, 1982 SFCS'08 23rd Annual Symposium on; 1982: IEEE.

[9] Cramer R, Damgård I, Nielsen J. Multiparty computation from threshold homomorphic encryption. Advances in cryptology—EUROCRYPT 2001. 2001:280-300.

[10] Ben-Or M, Goldwasser S, Wigderson A, editors. Completeness theorems for non-cryptographic fault-tolerant distributed computation. Proceedings of the twentieth annual ACM symposium on Theory of computing; 1988: ACM.

[11] Shamir A. How to share a secret. Communications of the ACM. 1979;22(11):612-3.

[12] Damgård I, Keller M, Larraia E, Pastro V, Scholl P, Smart NP, editors. Practical covertly secure MPC for dishonest majority–or: breaking the SPDZ limits. European Symposium on Research in Computer Security; 2013: Springer.

[13] 濱田 浩気, 木村 映善, 菊池 亮, 千田 浩司, 岡本 和也, 真鍋 史朗, et al. 秘密計算による分散医療統計システムの実装評価 (情報通信マネジメント). 電子情報通信学会技術研究報告. 2016 2016/05/26;116(65):111-7.

[14] Eizen KIMURA, Koki HAMADA, Ryo KIKUCHI, Koji CHIDA, Kazuya OKAMOTO, Shirou MANABE, et al. Evaluation of Secure Computation in a Distributed Healthcare Setting. Studies in health technology and informatics. 2016 2016/08;228:152-6.

[15] 星野伸明. 公的統計ミクロデータ提供制度の課題. 日本統計学会誌 シリーズ J= Journal of the Japan Statistical Society Japanese issue. 2010;40(1):23-45.

[16] Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Nordholt ES, Spicer K, et al. Statistical disclosure control: John Wiley & Sons; 2012.