部会資料

仮名加工情報等作成時の仮名加工技術の考察

令和 6 年 12 月

日 本 製 薬 工 業 協 会

医薬品評価委員会 データサイエンス部会

2023 年度　継続タスクフォース 2

目次

1. はじめに

新たな医薬品を開発するためには，人での有効性と安全性を評価するための臨床試験が必要となる。一般に臨床試験は少数例の試験から開始され，段階的に例数を増やしながら承認申請まで数試験から十数試験が行われる。一連の臨床試験の実施には膨大なリソース（ヒト，モノ，カネ，情報）が必要であり，また一つの試験に数か月から年単位の時間がかかる。このため，画期的な新医薬品をより早く医療現場に届けるためには臨床試験に代わる新たな有効性および安全性の評価方法を確立する必要があるという指摘がしばしばなされてきた。例えば，2004 年には米国食品医薬品局（Food and Drug Association，以下 FDA）が医薬品の人での評価を臨床試験だけに頼っていては医療技術の開発に時間とお金がかかりすぎてしまい，科学の進歩に対して新たな医療技術の提供が追い付かないと警鐘を鳴らしている [1)]。長年にわたりこのような指摘がなされているものの，残念ながら今日においてもこの課題が十分に解決したとは言い難い。

近年，国内外のリアルワールドデータ（RWD）や過去に実施された類薬の臨床試験データを二次利用して外部対照群のデータを構成することや，それらのデータから得られた知見を活用してより精度の高い臨床試験をデザインしようという試みが盛んに行われるようになってきた。これらの手法は未だ臨床試験に置き換わるほどの有効性・安全性の評価方法には発展していないものの，臨床試験の規模の縮小や期間の短縮には十分貢献し得る方法として大きな注目を浴びており，産官学で様々な取組みがなされている。例えば，世界の大手医薬品企業が加盟している非営利団体の TransCelerate Biopharm Inc.では，臨床試験デザインの改善，臨床試験実施の迅速化，疾患のより良い理解などを目的として，加盟企業が過去の臨床試験データを共有できるプラットフォーム（The Historical Trial Data Sharing initiative[1]）を構築している。また，欧州を中心とした国際的ながん研究ネットワークの ARCAD（Aide et Recherche en Cancérologie Digestive）は，がん治療に関する臨床試験データを共有して研究者間の協力を促進し，がん治療の向上を目指している。この ARCAD と連携して同様の活動をアジアに展開しているのが，国立がん研究センター東病院が中心となって設立された ARCAD アジア [2]である。ARCAD アジアでは，国立がん研究センター内にデータセンターを設置して，アジアを中心に実施された過去の臨床試験や臨床研究で得られたデータを統合したデータベースを構築し，単一の試験や研究では検討できないような研究課題へのアプローチや医薬品の研究開発活動の効率化などの実現を目指している。政府も 2013 年および 2022 年（閣議決定）の内閣府の健康・医療戦略において，健康・医療データ利活用の促進とデータ利活用基盤の構築を掲げており，2018 年には次世代医療基盤法 [3]を施行し，個人の医療情報を匿名加工することで，それらの情報を個人の同意を得ることなしに医療分野の研究開発に活用できるように法整備を進めた。さらに，同法の令和 5 年の改正（改正次世代医療基盤法）では，匿名加工よりも加工基準が緩やかな仮名加工医療情報 [4]を利活用する仕組みが創設され，医療情報がより広い範囲の研究に利用できるようになった。仮名加工は匿名加工に比べて加工方法が容易なだけでなく，データの加工による情報の損失が少ないため，データの有用性が向上することが

[1] https://www.transceleratebiopharmainc.com/initiatives/historical-trial-data-sharing/
[2] https://www.ncc.go.jp/jp/ncce/division/arcadasia/index.html
[3] https://www8.cao.go.jp/iryou/gaiyou/gaiyou.html
[4] 医療情報を，その区分に応じて掲げられた措置を講じて他の情報と照合しない限り特定の個人を識別することができないように加工して得られる個人に関する情報

期待されている。

改正次世代医療基盤法では，国が認定した事業者（認定仮名加工医療情報作成事業者，以下，認定作成事業者）が仮名加工医療情報を作成し，それを国が認定した利用者（認定利用事業者）のみが利活用できるという制度設計になっている。このため，製薬企業が仮名加工医療情報を利用する場合には，認定利用事業者として国から認定を受け，認定作成事業者から仮名加工医療情報を入手することになると思われる。従って，製薬企業が自ら仮名加工医療情報を作成することは想定しにくい。しかしながら，仮名加工医療情報の利用者として仮名加工技術に関する基本的な知識は必要であろう。仮名化を実現するためのデータの加工方法を知ることで，仮名加工医療情報の活用可能な範囲を知ることができるだけでなく，仮名加工医療情報の活用方法に応じた適切な仮名加工の方法を認定作成事業者とともに検討することも可能になると思われる。

そこで本報告書では，改正次世代医療基盤法の概略と仮名加工医療情報作成時の仮名加工方法について海外の報告書に基づき考察した。ぜひ，本報告書を一読いただき，今後の仮名加工情報活用の一助になれば幸いである。

2. 次世代医療基盤法

次世代医療基盤法（正式名称：医療分野の研究開発に資するための匿名加工医療情報に関する法律）は，医療情報を匿名加工し医療分野の研究開発での活用を促進するための個人情報保護法（正式名称：個人情報の保護に関する法律）の特別法として 2017 年に制定され，2018 年より施行されている。

附則に定められた施行後 5 年の見直しにより匿名加工医療情報では対応できない研究現場のニーズの対応がなされた。具体的には①希少な症例についてのデータ提供，②同一対象群に関する継続的・発展的なデータ提供，③薬事目的利用の前提であるデータの真正性確保するための元データに立ち返った検証を可能にする仮名加工医療情報の創設である（2023 年 5 月 26 日公布，2024 年 4 月 1 日施行）。この際，正式名称も医療分野の研究開発に資するための匿名加工医療情報及び仮名加工医療情報に関する法律（以後，改正次世代医療基盤法）に変更されている。

医薬産業政策研究所からの提言[5]にもある通り，仮名加工医療情報の新設により，前述の研究現場のニーズが充足され，利活用が促進されることが期待されている。なお，匿名加工医療情報と仮名加工医療情報のイメージについては内閣府より図 1 の通り例示されており，仮名加工医療情報は，特異な検査値や病名であっても削除・改変は原則不要とされている。

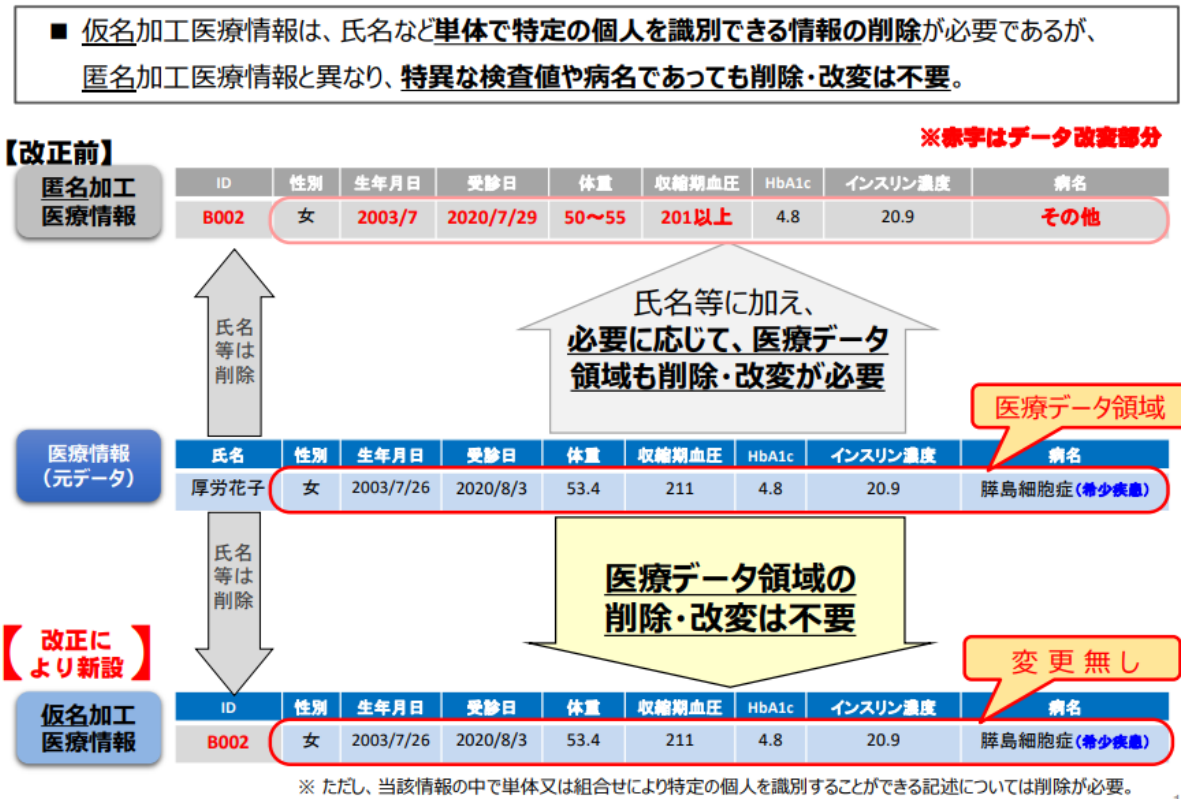


図 1 匿名加工医療情報と仮名加工医療情報の違い（イメージ）[6]

[5] 次世代医療基盤法がより良い制度となるために
https://www.jpma.or.jp/opir/news/070/k9rmj200000006e8-att/70_2.pdf

[6] https://www8.cao.go.jp/iryou/kouhou/pdf/kaisei_jisedaiiryou_rikatsuyou.pdf

2.1　改正次世代医療基盤法と個人情報保護法の違い

改正次世代医療基盤法と個人情報保護法の用語や定義の違いを表 1 にまとめる。

個人情報は,「生存する個人に関する情報」と規定されているのに対して，医療情報は該当箇所が無いため，故人に関する情報が含まれる。

医療情報は個人情報保護法では要配慮個人情報として位置づけられるため，法第 27 条第 2 項の規定による第三者提供（オプトアウト [7]による第三者提供）が認められないが，改正次世代医療基盤法ではあらかじめ本人に通知 [8]することを条件にオプトアウトによる認定作成事業者への医療情報の提供が認められている。

表 1 個人情報保護法と改正次世代医療基盤法の定義まとめ

	個人情報保護法（法と記載）	改正次世代医療基盤法（次法と記載）
XX 情報	個人情報（法 2①） 生存する個人に関する情報であって，次の各号のいずれかに該当するものをいう。 一　当該情報に含まれる氏名，生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ，それにより特定の個人を識別することができることとなるものを含む。） 二　個人識別符号が含まれるもの 要配慮個人情報（法 2③） 本人の人種，信条，社会的身分，病歴，犯罪の経歴，犯罪により害を被った事実その他本人に対する不当な差別，偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものとして政令で定める記述等が含まれる個人情報をいう。	医療情報（次法 2①） 特定の個人の病歴その他の当該個人の心身の状態に関する情報であって，当該心身の状態を理由とする当該個人又はその子孫に対する不当な差別，偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものとして政令で定める記述等であるものが含まれる個人に関する情報のうち，次の各号のいずれかに該当するものをいう。 一　当該情報に含まれる氏名，生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ，それにより特定の個人を識別することができることとなるものを含む。） 二　個人識別符号が含まれるもの

[7] オプトアウト：あらかじめ個人情報保護委員会（主務大臣：次法の場合）に届け出ること及び本人に通知する（又は本人が容易に知り得る状態に置く：次法では認められない規程）ことで事前の同意を得ることなく第三者提供を行い，本人（又はその遺族から：次法のみの追加規程）の求めに応じて提供を停止する方法

[8] 本人に対する通知：医療情報取扱事業者が本人に対する通知を実施するに至った以降での最初の受診時に書面を交付する方法を基本として，医療情報取扱事業者ごとに適切な方法を選択する必要がある。（次法ガイドライン V）

	個人情報保護法（法と記載）	改正次世代医療基盤法（次法と記載）
第三者提供	法律上の例外規定[9]を除き，あらかじめ本人の同意を得ないで，個人データを第三者に提供してはならない。（法 27①） 第三者に提供される個人データが要配慮個人情報である場合は，オプトアウトによる第三者提供が認められない。（法 27②）	あらかじめ，本人に通知[8]するとともに，主務大臣に届け出たときは，当該医療情報を認定匿名（仮名）加工医療情報作成事業者に提供することができる。（次法 52①）（次法 57①）
匿名 XX	匿名加工情報 特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報（法 2⑥） 個人情報ではなくなるため，本人同意を得なくとも，第三者提供や目的外利用が可能	匿名加工医療情報 特定の個人を識別することができないように医療情報を加工して得られる個人に関する情報であって，当該医療情報を復元することができないようにしたもの（次法 2③） 国から認定された認定匿名加工医療情報作成事業者のみが作成可能
仮名 XX	仮名加工情報 他の情報と照合しない限り特定の個人を識別することができないように加工された個人に関する情報（法 2⑤） ※対照表と照合すれば本人が分かる程度まで加工されたもの（個人情報に該当） 特定の条件を満たせば目的外利用が可能だが，第三者提供は不可	仮名加工医療情報 他の情報と照合しない限り特定の個人を識別することができないように医療情報を加工して得られる個人に関する情報（次法 2④） 個人情報から氏名や ID 等の削除が必要だが，匿名加工医療情報とは異なり，特異な値や希少疾患名等の削除等は不要。 国から認定された認定仮名加工医療情報作成事業者のみが作成可能で，利用には仮名加工医療情報利活用者の認定が必要

2.2 仮名加工に注目した海外での定義との比較

2.2.1 EU や ISO における仮名化（Pseudonymisation）の定義

EU の一般データ保護規則（General Data Protection Regulation，以下 GDPR）で仮名化は次のよう

[9] 例外規定
①法令に基づく場合
②人の生命，身体又は財産の保護のために必要がある場合であって，本人の同意を得ることが困難であるとき
③公衆衛生の向上又は児童の健全な育成の推進のために特に必要がある場合であって，本人の同意を得ることが困難であるとき
④国の機関若しくは地方公共団体又はその委託を受けた者が法令の定める事務を遂行することに対して協力する必要がある場合であって，本人の同意を得ることにより当該事務の遂行に支障を及ぼすおそれがあるとき

に定義されている。

「仮名化」とは，追加的な情報[10]が分離して保管されており，かつ，その個人データが識別された自然人[11]又は識別可能な自然人に属することを示さないことを確保するための技術上及び組織上の措置の下にあることを条件として，その追加的な情報の利用なしには，その個人データが特定のデータ主体[11]に属することを示すことができないようにする態様で行われる個人データの取扱いを意味する。（GDPR 第 4 条⑤）

仮名化はデザインによるデータ保護（data protection by design）[12]を実現させるための技術（GDPR 第 25 条），個人データ取扱いの安全性を確保する手段（GDPR 第 32 条）として明示的に言及されており，公共の利益における保管の目的，科学調査もしくは歴史調査の目的または統計の目的のための取扱の保護措置としても例示されている（GDPR 第 89 条）

また，ISO 25237:2017 では次のように定義されている。

‘particular type of de-identification that both removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms’ [ISO，2017]

（執筆者訳：データ主体との関連付けを削除し，データ主体に関連する特定の特性の組み合わせと pseudonym（仮名）との関連付けを追加する，非特定化の手法の一つ）

2.2.2　個人情報保護法と GDPR や ISO の仮名化の違い

日本の個人情報保護法における仮名加工情報は個人情報の種類の一つとして定義されている。一方，GDPR や ISO における仮名化は個人データの安全性を高める手段（取扱い方法）の一つとして定義されている。

[10] 追加的な情報とは個人情報保護法の仮名加工情報の「他の情報と照合しない限り」の「他の情報」に該当する GDPR 翻訳上の用語

[11] 自然人とは法律用語で法人と対比される概念で「人」や「個人」と表記される。GDPR ではこの識別され得る自然人のことを「データ主体（data subject）」と呼ぶ

[12] 製品やサービスの企画・設計段階からの個人データ保護

3. 仮名加工医療情報の作成技術について

医療情報から仮名加工医療情報を作成するためには，当該医療情報に含まれる「氏名，生年月日その他の記述等により特定の個人を識別できるもの」[13]の一部を削除することおよび個人識別符号が含まれていればそれらを削除する必要がある。なお，ここでの「削除」には，「当該一部の記述等」又は「当該個人識別符号」を「復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む」とされており，個人 ID など医療情報の利活用に必要であれば削除せずに特定の個人を識別できないように復元不能な値や記述に置き換えても良いことになっている。ただ，残念ながら国内には具体的な仮名加工方法を示したガイダンスや文献は現時点で存在しておらず，どのような方法で個人を特定できない情報に置き換えればよいのかという疑問が残る。このため，本報告書では欧州連合サイバーセキュリティ機関（European Union Agency for Cybersecurity, 略称 ENISA）が公表しているレポート（表 2 参照）を参考に仮名加工医療情報の作成方法を考察した。3.1 章では仮名加工情報の設計目標，3.2 章では製薬企業における仮名加工医療情報の活用場面, 3.3 章では仮名化ポリシー, 3.4 章では仮名加工情報の作成技術を整理する。

表 2 本報告書が参考にした ENISA の公表レポート

発行時期	タイトル	概要
2018 年 11 月	Recommendations on shaping technology according to GDPR provisions An overview on data pseudonymisation[2)]	標準的な仮名加工情報作成技術と設計目標を紹介
2019 年 11 月	Pseudonymisation techniques and best practices Recommendations on shaping technology according to data protection and privacy provisions[3)]	仮名加工情報が利用される複数のシナリオを紹介。併せて，様々なタイプの再識別化攻撃と，それらの攻撃に対する仮名化技術とポリシーを紹介
2021 年 1 月	Data Pseudonymisation: Advanced Techniques and Use Cases Technical analysis of cybersecurity measures in data protection and privacy[4)]	発展的な仮名加工情報作成技術とヘルスケア分野における具体的な使用事例を紹介
2022 年 3 月	Deploying Pseudonymisation Techniques The case of the Health Sector[5)]	医療分野における仮名加工情報利用のユースケースを紹介
2023 年 1 月	Engineering Personal Data Sharing Emerging Use Cases and Technologies[6)]	仮名加工情報を共有する際の注意点などを紹介

3.1 仮名加工情報の設計目標（Design Goals）について

本章の冒頭でも述べたように仮名加工する際の重要な点は，データ主体に関連する識別情報

[13] 他の情報と容易に照合することができ，それにより特定の個人を識別することができることとなるものを含む

（personal identifiers，個人識別子）を削除または仮名に置き換えることによって「隠す」ことである。再識別を可能にするために個人識別情報と pseudonym（仮名）の間の関連性を保持しておく必要がある場合には，仮名加工の実施者が適切な管理下でその情報を維持しておく。

仮名化をする際に，個人情報の保護を強化するとデータの有用性が低くなり，逆にデータの有用性を維持しようとすると個人情報保護に対するリスクが増えるというトレードオフの関係があるため，個人情報の保護をどこまで強めるのか，あらかじめきちんと計画（設計）する必要がある。仮名加工情報を作成するにあたり，仮名加工情報作成者は特定のデータ処理作業が個人の権利と自由に及ぼすリスクを考慮したうえで, 最適な手法の採用に向けて以下の設計目標（Design Goals）を設定できる。

- 設計目標 1) 第三者が pseudonyms から容易に再特定化（re-identification）ができないこと
- 設計目標 2) 第三者が pseudonyms を再生成するのは容易ではないこと（異なるデータ処理領域間で，同じ pseudonyms が使用されることを避ける（領域間の非連結性を確保する）ため）

これらの設計目標はデータ管理者が「追加的な情報」にアクセスして仮名化プロセス後にデータ主体を再特定できるという前提に基づいている。

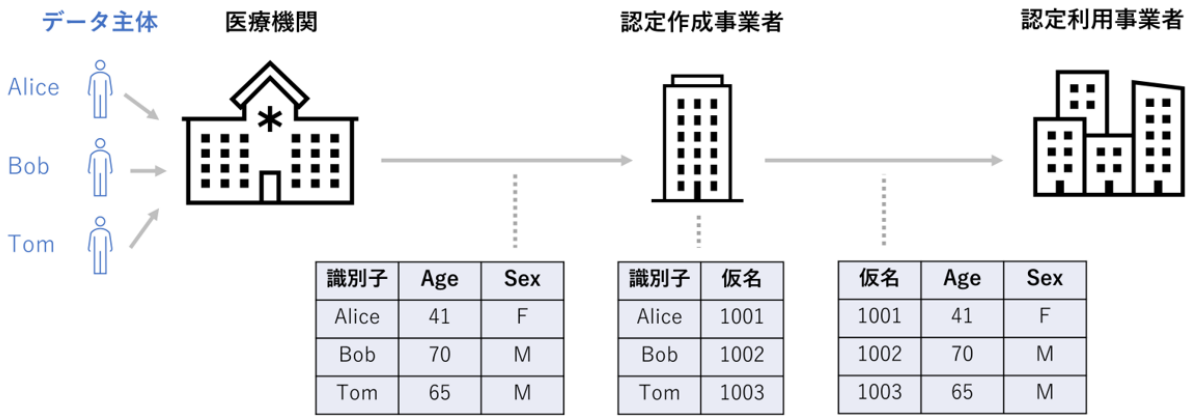
なお，データ管理者が仮名化されたデータを特定の初期識別子（initial identifier）と関連付ける必要がない場合もある。例えば，管理者は個人の追跡（tracking）を行うだけでよい場合がある。すなわち，実際には初期識別子の情報がなくとも，特定の処理状況内で個人を他の個人から区別することができるだけで十分な状況もあり，同じ pseudonym が常に同じ個人に割り当てられることを保証する仮名化技術を採用することで，このような要件を満たす手段となり得る。

また，仮名化にはデータの正確性の観点から追加のデータ保護の利点をもたらす可能性があり，第 3 の設計目標となる。例えば，初期識別子から数学的なアルゴリズムで pseudonym を生成する仮名化技術が存在するため，これらの pseudonym は特定の状況下でデータ主体の ID を検証するのに十分であると考えられる。なぜなら，データ主体の ID を明かすことなく，pseudonym が特定の ID に対応していることが確認できるからである。

このように状況にあった設計目標をたてることによって，適切な仮名加工技術の選択を可能とする。

3.2 製薬企業による仮名加工医療情報の活用場面

改正次世代医療基盤法の下では，認定仮名加工医療情報作成事業者（以下，認定作成事業者）が仮名加工医療情報を作成し，認定仮名加工医療情報利用事業者（以下，認定利用事業者）のみがそれを利用できる。従って，ここでは改正次世代医療基盤法の下，製薬企業が認定利用事業者となり認定作成事業者から薬事申請やエビデンス創出の目的で仮名加工医療情報を入手するシナリオを想定する。



本シナリオでは，データ主体である患者の情報を医療機関が入手し，認定作成事業者が仮名加工処理を実施したのち，そのデータを製薬企業が入手する状況を想定している。なお，仮名加工医療情報に対する安全管理措置に関して，認定利用事業者が自ら整備した環境下に仮名加工医療情報を保存することが可能な I 型認定と，認定作成事業者等が整備した Visiting 環境での利用に限定する II 型認定の 2 種類が存在し，認定利用事業者はいずれかの認定を取得して仮名加工医療情報を利用する状況を想定する。

匿名加工医療情報であれば再特定のリスクをなくすために，年齢のカテゴリ化や頻度の少ないデータの削除などの加工が実施される。一方で，仮名加工医療情報ではそのような加工は想定されず，ID 情報のみを仮名化したデータセットを入手することになる。

ただし，薬事申請を目的に仮名加工医療情報を利用する場合には，各国規制当局の適合性調査やデータの品質担保の目的で，原資料に戻るために仮名化処理した ID 情報を元の ID 情報に復元する必要がでてくる（データの**再識別**）。

さらに，ヘルスケア領域のデータの特性として，同一被験者が複数のデータセットに含まれることがある（例えば被験者背景，有効性，安全性のデータ項目ごとにデータセット化される）。そのため，仮名化した ID 情報を用いて複数データセットにまたがる同一患者を特定する必要があるため，**追跡性（tracking）**の要件も満たす必要がある。

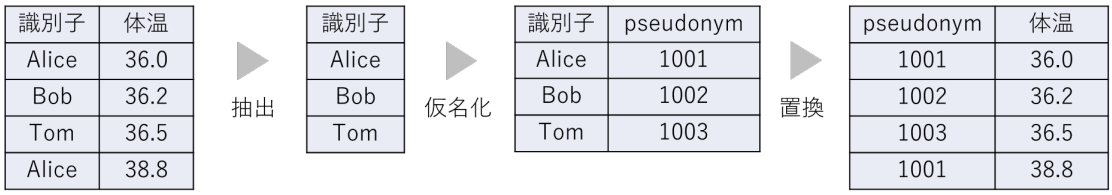
上記 2 点の特性（再識別，追跡性）に注目して，3.4 章では仮名加工医療情報の作成技術を整理していく。

3.3　仮名化ポリシーについて

次章で解説する仮名加工技術の選択は重要であるが，仮名化の実装のポリシーもその実用的な応用において同様に重要である。本章では，複数の識別子を含むデータベースや任意の文書の仮名化といった，より一般的な問題を考察する。2 つのデータセットAとBに何度も現れる識別子を考えてみよう。仮名化後，識別子は次の 3 つのポリシーのいずれかに従って置き換えられる：①決定的仮名化（Deterministic pseudonymisation），②文書ランダム化仮名化（Document-randomized pseudonymisation），③完全ランダム化仮名化（Fully-randomized pseudonymisation）。それぞれのポリシーについて以降解説していく。なお，前章で述べた製薬企業による仮名加工医療情報の活用場面を考えると，①決定的仮名化のポリシーを適用することが想定される。

3.3.1　決定的仮名化（Deterministic pseudonymisation）

識別子が，すべてのデータベースに対して，その識別子が出現するたびに，常に同じ pseudonym に置き換えられる。pseudonym はデータベース内や異なるデータベース間で一貫性がある。この方針を実装するために，まずデータベースに含まれるユニークな識別子のリストを抽出する。次に，このリストを pseudonym にマッピングし，最終的にデータベースの識別子を pseudonym に置き換える。



3.3.2　文書ランダム化仮名化（Document-randomized pseudonymisation）

識別子が 1 つのデータベースに現れるたびに，それは異なる pseudonym（$pseudo1, pseudo2,$ …）に置き換えられる。しかし，識別子は常に複数のデータセットAおよびBの中では同じ pseudonym の集合（$pseudo1$, $pseudo2$）にマッピングされる。例えば，図 2 で示した例で Alice は出現する度に pseudonym の集合（1032，1982，1547，…）から順番に pseudonym に置き換えられる。

この場合，仮名化は異なるデータベース間でのみ一貫性がある。識別子と仮名の対応表は，データベースに含まれるすべての識別子を使用して作成される。特定の識別子（例：Alice）の各出現は独立して扱われる。

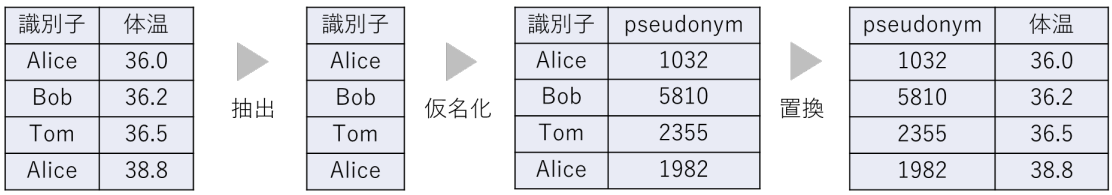
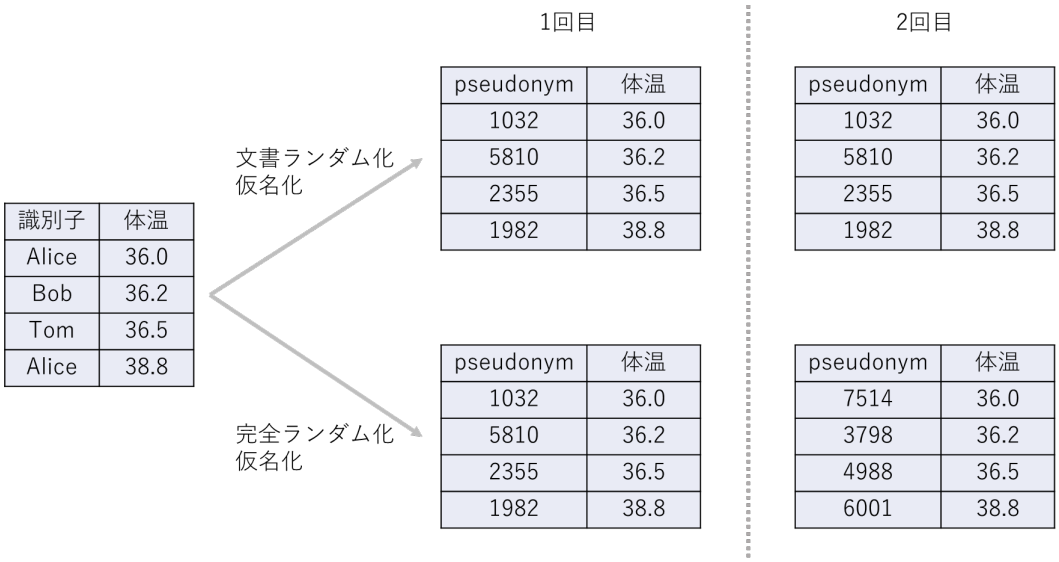


図 2　文書ランダム化仮名化の例

3.3.3　完全ランダム化仮名化（Fully-randomized pseudonymisation）

データベースAまたはB内の識別子の任意の出現に対して，識別子は異なる pseudonym（$pseudo1$，$pseudo2$）に置き換えられる。このポリシーは，文書ランダム化仮名化のさらなる拡張と見なすことができ，両者は単一のデータに適用される場合は同様の結果となる。しかし，同じデータに完全ランダム化仮名化を 2 回行うと，1 回目と 2 回目で異なる結果が得られる。一方，文書ランダム化仮名化では，同じ結果が 2 回得られる。つまり，文書ランダム化仮名化ではランダム性が選択的（例えば，Alice のみ）であるのに対し，完全ランダム化仮名化ではランダム性が任意のレコードに適用されるため全体的である。



3.4　仮名加工技術について

仮名加工技術は単純なものから複雑なものまで複数の手法が存在する。代表的な手法と各手法を用いた際の再識別の方法を表 3 に整理する。

表 3　代表的な仮名加工技術と再識別の方法

仮名加工技術の種類	仮名化の例	再識別（Recovery）の方法
Counter/ **カウンター**	連番で ID を置き換える **1，2，3，…**	対応表
Random number/ **乱数**	0000～9999 のユニークかつランダムな数字で ID を置き換える **9170，3069，1415，…**	対応表
Cryptographic Hash function/ **暗号学的ハッシュ化**	John をハッシュ化 [14] **a8cfcd74832004951b4408cdb0a5dbcd8c7e52d43f7fe244bf720582e05241da**	対応表
Hash-based Message Authentication Code（HMAC）/ **秘密鍵を使った暗号学的ハッシュ化**	John を秘密鍵”1337”でハッシュ化 [14] **602786317db1ba3396fbf3dfcaa34bb400c4c4778b863b2bc66bd14dcc19b95f**	対応表

[14] ハッシュ関数のうち SHA256 を使用。SHA256 は SHA-2（Secure Hash Algorithm 2）に分類され，256 ビットの 16 進数のハッシュ値を生成する暗号的ハッシュ関数である。

仮名加工技術の種類	仮名化の例	再識別（Recovery）の方法
Encryption/ 暗号化	John を秘密鍵”1337”で暗号化 [15] **msPwj/67G0bkJgIjRJgK9w==**	秘密鍵（Decryption/ 復号化）

ENISA レポート（2019 年 11 月）の Table 4 および ENISA レポート（2022 年 3 月）の Figure 3 を参照した

最も簡単な仮名加工技術としてカウンターや乱数があるが，いずれの手法も対応表がない限りは追跡ができない。大規模なデータベースの場合，すべての ID に対応する pseudonym の対応表を作成し管理することはほぼ不可能である。つまり，これらの簡単な手法は大規模かつ複雑なデータセットに対しては拡張性（Scalability）の観点で課題がある。また，乱数には衝突（異なる ID に対して同一の pseudonym が割り当てられること）のリスクもある。したがって，想定する仮名加工医療情報の利用場面（薬事承認を目的とした大規模な臨床データの活用）では，3.3.1 章の①決定的仮名化のポリシーを適用してデータベース内およびデータベース間で pseudonym の一貫性を担保する必要があるため，カウンターや乱数は実用性に乏しい。

一方，ハッシュ化はオリジナルの ID にハッシュ関数をかけることで固定長のハッシュ値に変換するという手法である。暗号学的ハッシュ関数 h は，任意の長さの入力メッセージ m を固定サイズの出力 h(m)（例：サイズ 256 ビット，すなわち 32 文字）に変換する特定の特性を持つ関数である。出力 h(m)はハッシュ値あるいはメッセージ・ダイジェストと呼ばれる。

暗号学的ハッシュ関数により生成されるハッシュ値は次の特性を満たす。

性質 i)　h(m)がわかっても m を計算上求めることができない

性質 ii)　与えられた m に対し，h(m’)=h(m)となる別の m’を計算上求めることができない

性質 iii)　h(m’)=h(m)となる m と m’を計算上求めることができない

手法の特性上，対応表がない限りはオリジナルの ID の再識別はできないが，同一患者に対して単一の仮名化ができるため追跡が可能である。なお，オリジナルの ID と仮名化された ID の対応表を作成事業者側で保有しておくことで，薬事申請などの際にオリジナル ID との連結が可能となり，対応表と併用することで再識別も可能である。

本手法の欠点として，同じ識別子に同じ暗号学的ハッシュ関数を適用する第三者が同じ仮名を取得するため，設計目標 2）が保持されない（第三者でも同じ m と暗号学的ハッシュ関数を使えば同じ h(m)を生成できる）点があげられる。

なお，入力となるオリジナルの ID は任意の長さであるのに対して，出力のハッシュ値は固定長であるため，「同じハッシュ値となる異なる平文の組み合わせ」が理論上存在する。つまり「衝突」

[15] 暗号化手法のうち AES を使用。AES（Advanced Encryption Standard）は，共通鍵暗号方式の一つで，同じ鍵を使用して暗号化と復号を行う。データをブロックに分割し，複雑な数学的計算を行い，鍵を使用して暗号化する。

のリスクがある。衝突を回避する方法[16]として，秘密鍵を併用する HMAC（Hash-based message authentication code）や，salt を併用する Salted hash function などの手法が存在する。HMAC では，オリジナルの ID に予測不能で十分な長さの秘密鍵を連結した上でハッシュ化する方法である。また，Salted hash function は「salt」と呼ばれる補助的なランダムな文字列を追加することによって入力の文字列に手を加える手法である。

前出の（秘密鍵や salt を用いない）暗号学的ハッシュ関数との主な違いは，同じ入力に対して，特定の鍵の選択に従っていくつかの異なる仮名を生成できることであり，従って，設計目標 2）が保証される。さらに，第三者（例えば敵対者）が秘密鍵を知らない限り設計目標 1）も保持される。なお，データ管理者が同じ個人に同じ仮名を割り当てる必要がある場合は，同じ秘密鍵を使用しなければならない。

暗号化（Encryption）も，pseudonym を得るための効率的な方法である。オリジナル ID は，暗号化アルゴリズム（例 : 暗号化標準である AES［FIPS，2001］）を用いて暗号化することができ，これにより pseudonym が得られる。同じ秘密鍵を共有することで復号化（Decryption）が可能となり，仮名化した ID からオリジナル ID の再識別が可能となる。

仮名加工医療情報を活用する場面では，改正次世代医療基盤法にて安全管理措置を講ずることが義務付けられている。そのため，安全な環境下でデータ利用をすることから外部からのブルートフォース攻撃（総当たり攻撃）などは想定されず，設計目標 1)と 2)を満たした上で再識別と追跡性の両側面の機能担保が必要となると想定される。その場合に利用される仮名加工技術としては以下の 2 通りの方法が考えられる。

- 暗号学的ハッシュ化（HMAC 含む）および対応表の維持管理
- 暗号化および秘密鍵の維持管理

2024 年 4 月に次世代医療基盤法ガイドラインが発出され仮名加工医療情報の加工の基準が示されたが，仮名加工技術としての言及は「他の記述等に置き換える場合は，元の記述等を復元できる規則性を有しない方法でなければならない」との記載のみで，技術的手法には言及していない。また，「認定仮名加工医療情報利用事業者における医療分野の研究開発のために必要でない情報については，追加的な削除又は加工を行うことが望ましい」との記載もあり，認定利用事業者ごと，もしくは研究内容ごとに仮名加工処理が変わる可能性もある。このため，製薬企業が認定利用事業者として仮名加工医療情報を受け取る際にはどのような仮名加工技術が利用されているかを認定作成事業者と事前にすり合わせ，データの加工に対する共通理解を持つことが望まれる。

[16] 異なる入力値に対してハッシュ値が絶対に衝突しない（同じ値にならない）方法ではない。ハッシュ値が一致してしまう異なる入力値を見つけることが極めて困難な方法という意味である。A hash function cannot be free of collisions because it is a compression function with a set hash length. The collision-free condition simply indicates that these collisions should be difficult to locate. Cryptography - Hash functions (tutorialspoint.com)から引用

4. おわりに

本報告書では，認定利用事業者が認定作成事業者から薬事申請目的で仮名加工医療情報を入手する状況を想定し，その際に必要となる仮名加工医療情報の作成技術について考察した。次世代医療基盤法の令和 5 年の改正によって仮名加工医療情報を利活用する仕組みが創設されたものの，具体的な仮名加工医療情報の作成ルールについては現時点で共通認識がなく，今後運用されながら方法論の議論・整理が進んでいくものと思われる。薬事申請目的での仮名加工医療情報の利用を想定した際は，製薬企業自身が仮名加工医療情報を作成することはないと思われるが，仮名加工医療情報の利用者として仮名加工のプロセスや仮名加工されたデータの限界を理解するためにも一定の知識を持っておく必要があると考える。また，企業内の医療データの二次利用の場面においても，データのセキュリティ確保のために仮名加工が必要な状況も想定されるため，3.4 章で紹介した各種手法の特徴を踏まえ，データの特徴や利用目的に応じて適切な仮名加工データの作成技術を検討することが望まれる。仮名化は，単にデータ内に存在する識別子を別の値に置き換えればよいというわけではなく，個人情報の保護とデータの有用性の維持という相反する目的を実行可能な範囲で両立させるために考慮すべき点が多くあることを理解いただけたなら幸いである。

5. 資料作成者

医薬品評価委員会 データサイエンス部会 2023 年度 継続タスクフォース 2

加藤　智子	サノフィ株式会社（担当副部会長）
小山　暢之	第一三共株式会社（TF リーダー）
青木　真	アステラス製薬株式会社（TF リーダー）
田嶋　幸聖	中外製薬株式会社（TF リーダー）
佐土原　和宏	鳥居薬品株式会社
大塚　晶仁	ゼリア新薬工業株式会社
吉岡　彬生	富士フイルム富山化学株式会社（～2024 年 3 月）

6. 参考文献

1) Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products Report. FDA, 2004
2) Recommendations on shaping technology according to GDPR provisions. An overview on data pseudonymisation, ENISA, 2018
3) Pseudonymisation techniques and best practices. Recommendations on shaping technology according to data protection and privacy provisions, ENISA, 2019
4) Data pseudonymisation. Advanced techniques and use cases: Technical analysis of cybersecurity measures in data protection and privacy, ENISA, 2021
5) Deploying pseudonymisation techniques. The case of the health sector, ENISA, 2022
6) Engineering personal data sharing. Emerging use cases and technologies, ENISA, 2023