

令和 6 年度厚生労働科学研究費補助金
(循環器疾患・糖尿病等生活習慣病対策総合研究事業)

都市・農村における生活習慣病の実態比較およびパーソナルヘルスレコードを
活用した重症化予防介入プログラムの開発と効果検証
分担研究報告書【4】

生活習慣病の発症および重症化予防介入としての PHR 活用にかかる課題の検討

<研究分担者>

山本景一 大阪歯科大学医療イノベーション研究推進機構
事業化研究推進センター データサイエンス部門・教授

研究要旨

パーソナルヘルスレコード (PHR) による生活習慣病の発症・重症化予防介入プログラムの開発にあたり、日々更新されるライフログデータ、お薬手帳、その他さまざまなデータを統合し、分析を行わなければならない。本分担研究では、PHR による収集データと健康医療介護統合データベースからの解析用データの統合手法を確立する。特にウェアラブルデバイスによる頻回計測データの分析手法は世界的に未だ確立した方法論は存在しない。頻回計測データの分析とそれを活用した生活習慣病の発症・重症化予防を実現するための PHR データ収集とデータベース化、およびデータ流通における標準化について、課題検討を行った。

Kuan らは心不全の状態を評価する血液検査値である NT-proBNP に臨床指標を加えることで診断性能を高める研究を報告している。また米国の Digital Medicine Society は、複数ソースから得られる健康に関する頻回計測センサーデータを統合利用するための Sensor Data Integration Project を実施している。我が国でも、「民間事業者の PHR サービスに関わるガイドライン (第 2 版)」(一般社団法人 PHR 普及推進協議会作成)にて、「医療機関-PHR 間、PHR-PHR 間、計測機器-PHR 間、スマホ OS 標準アプリ-PHR サービス間」の標準データ交換規格の案を提示している。加えて、Brokamp らは、DeGAUSS と呼ばれる生活環境に関する大規模時空間データベースを構築し、日次大気汚染物質暴露を推定するジオマーカー評価を行っている。特に、PHR データ標準規格として、Open mHealth (アプリやデバイス間でデータを共有し、医療機関等に提供する目的で開発されたデータ交換のためのデータ標準) が使用され始めている。また、合成患者データ (ヘルスケアデータから人工的に機械学習を用いて作成されたデータの総称) は、元となった患者レベルの個票データは推定できないことから、プライバシー保護の観点で有用性が高い技術と考えられる。加えて米国では、CDC が 2019 年末に Data Modernization Initiative (DMI) を開始し、COVID-19 対策として 10 億ドル超を投資して高度な疾病監視システムを構築した。また、米国慢性疾患協会の MENDS (Multi-state EHR based-Network for Disease Surveillance) は、全米の電子カルテネットワークを活用し、慢性疾患の疫学指標をリアルタイム監視するダッシュボードシステムを提供している。

このように、医療・健康ビッグデータや AI 技術の発展を背景に、多様なデータソースを組み合わせるデータ活用と社会基盤整備が世界中で進められており、本事業の役割の重要が示唆された。

A. 研究目的

PHR による生活習慣病の発症・重症化予防に関する介入プログラムの開発にあたり、バイタサインや行動履歴などのライフログデータ、お薬手帳、その他さまざまなデータを統合し、分析を行わなければならない。健康の社会的決定要因 (SDoH) とは、人々の誕生・成長・生活・就業・加齢において継続的に健康状態に影響を及ぼす要因である。

SDoH では、健康を決定する要因として医療的要因は 20%に過ぎず社会的経済的要因・物理的環境要因・健康に関わる行動が 80%を占めるとされ、健康寿命延伸のために病院外の健康データの利活用の必要性が高まっている。従来医療におけるデータ収集は、来院時に検査や問診を行い医師・看護師・クリニカル・リサーチ・コーディネーターなどの専門職が記録する形で行われてきた。ウェアラブル技術の発展により、睡眠・血圧・血糖・歩数その他の日々の健康データをモバイルセンサー (ウェアラブルデバイス) で収集することが可能となった。単一のデータだけから得られる知見は限られており、カルテや健診などの院内の診療・健康データに加え、このような多様な頻回計測センサーデータを組み合わせ活用することが期待される。

すでに心拍・睡眠その他多数の身体データを計測する機器やデバイスが提案されている。一般にセンサーから得られるデータ量は従来の数千～数万倍とされる。そのような複数ソースから得られる大量の多次元時系列データを統合し活用するためには、データの標準化、可視化・分析手法の確立、同意取得を

含む社会的ルール作り、その他多くの課題がある。デバイス開発・データ収集・解析・介入を単一の企業やグループで行うのではなく複数者 (社) で分業する未来が想定され、標準的なデータ交換規格による相互運用性の確立と、本人の意思でデータを集約・活用できる社会基盤の確立が喫緊の課題である。

本研究では、データ分析と結果の個人へのフィードバックを前提とした PHR データ収集とデータベース化、およびデータ流通における標準化について課題検討を行いたい。

B. 研究方法

出版済み論文の検索、国内外の関連学会やシンポジウム等に参加し、情報収集を行う。集められた情報を本研究で開発するアプリや調査項目に迅速に反映を行う。

(倫理面への配慮)

本研究は侵襲性のある介入はなく、ヒトゲノムの情報も利用しない。

C. 研究結果

以下に PHR やウェアラブルデバイスによる頻回計測センサーデータ他の健康関連データを用いた治療・健康増進に関するいくつかの事例を紹介する。

「デジタルバイオマーカー」とは、スマートフォンやウェアラブル機器などから得られる心拍、歩数、睡眠他の心理・行動データを用いて、病気の有無や治療による変化を客観的に可視化する指標である。医療 AI の進展もあり、世界中の研究グループによる研究開発競争が行われている。例えば Kuan らは心

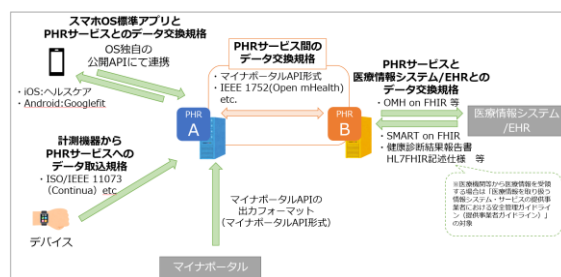
不全の状態を評価する血液検査値である NT-proBNP に年齢・推定糸球体濾過量

(eGFR)・ヘモグロビン・肥満指数

(BMI)・心拍数・血圧・末梢浮腫・慢性閉

塞性肺疾患や虚血性心疾患の既往などの臨床指標を加えることで診断性能を高める研究の報告を行っている [1]。

心拍・睡眠その他多数の身体データを計測する機器やデバイスが提案されている。マルチモーダル AI とは、数値/画像/テキスト/音声など複数種類のデータ（モダリティ）を組み合わせて処理できる単一の AI モデルであり、今後の発展が見込まれる。医療・健康分野でも PHR、Person Generated Data (PGD)、ウェアラブルデバイスなどの複数ソースから得られる大量の多次元時系列データを統合し活用することが期待されている。アメリカに本部を置く Digital Medicine Society (デジタルメディスン学会: DiMe)は、複数ソースから得られる健康に関する頻回計測センサーデータを統合利用するための Sensor Data Integration Project を実施し、データ生産者、データ処理者、データ利用者に対するユースケースとツールキットの開発を行っている[2]。我が国でも、一般社団法人 PHR 普及推進協議会が作成した「民間事業者の PHR サービスに関するガイドライン（第2版）」で、医療機関-PHR 間、PHR-PHR 間、計測機器-PHR 間、スマホ OS 標準アプリ-PHR サービス間の標準データ交換規格の案を提示している [3]。



【PHR 標準データ交換規格】

また、例えば大気汚染と喘息、近隣の犯罪と精神衛生、地域の緑地と IQ など、場所に基づく情報と健康上の課題との関係が調べられることがある。「ジオマーカー」と呼ばれる場所ベースの暴露は、健康の強力な決定要因であるが、従来オープンデータや統合ツールが少なく利用が困難であった。Brokamp らは、Decentralized Geomarker Assessment for Multi-Site Studies (DeGAUSS) [4]と呼ばれる生活環境に関する大規模時空間データベースを構築し、日次大気汚染物質暴露を推定するジオマーカー評価を行った事例他を公表している。

PHR のデータ標準としては、Open mHealth が利用され始めている。Open mHealth は患者自身でデータを生成、入力、アクセスを可能とすることを目標とし、アプリやデバイス間でデータを共有し、医療機関等に提供することを目的とするデータ交換のためのデータ標準である。アプリやデバイスでは無く、メタデータやデータに関する仕様であり、Apache 2 オープンソースライセンスの下、データ交換用 API ライブラリ、データ・リポジトリ、データ・ビジュアライゼーションツールなどが Github (<https://github.com/openmhealth>) で公開されている。一部は、IEEE 1752 で標準化されており、NIH Research Opportunity Announcement

OTA-21-015A (2021)で有望なデータ標準として引用されている。電子カルテを含む医療情報システム間のメッセージ交換規格である FHIR と異なるユースケースと異なる開発者コミュニティを対象としている。どちらもコンテンツモデルであり、互いにマッピング可能である。Open mHealth では既に約 90 種類の血圧、心拍数、血中グルコース、睡眠時間、身体活動量、メタデータ、他のスキーマ定義が公開されている。カリフォルニア大学サンフランシスコ校(UCSF Health)でモバイルヘルスの血圧検査値を電子カルテへのログインや書き込みなしでアクセス可能となっている[5]。

加えて、生成系 AI の技術を利用しカルテデータを生成する「合成患者データ (Synthetic Patient Data)」技術が世界的に注目をされ始めている。生成系 AI とは、画像生成、文章生成、音楽生成、図面生成などを含めた「0 から 1 を生み出す AI」のことであり、Chat GPT や Stable Diffusion の例がある。合成患者データとは、個人を特定できる情報は含まずに、ヘルスケアデータから人工的に機械学習などの技術を用いて作成されたデータの総称である。元となった患者レベルの個票データは推定できないため、プライバシー保護の観点で有用性が高い技術と考えられる。2019 年に実施されたアメリカの EHR COVID-19 DREAM Challenge プロジェクトにより世界的に注目を集め、関連する論文が増加した。診療データを研究利用する場合、患者の個人情報保護のために匿名加工が行われるが、匿名化・仮名化のみでは再特定化（再識別化）の可能性を排除できないためプライバシー保護は十分でない。匿名化することで情報が失われ、解析に必要な変数が使用でき

ないケースがある。匿名加工に代わる個人情報保護手法として、差分プライバシー（錯乱手法）がある。差分プライバシーとは、対象データにランダムな値（ノイズ）を追加することにより、個々のデータが特定されるリスクを最小限に抑えるための技術である。アメリカ国勢調査局（U.S. Census Bureau）は、2020 年人口国勢調査における統計表の作成・公表において差分プライバシーの方法論の適用が行われている。しかし、差分プライバシーはノイズによっては有用性が低下する可能性がある。よって次世代の個人情報保護手法として、各種公的統計、機械学習の教師データ生成、臨床研究の外部対照群、マイクロシミュレーション他に合成患者データの利活用が期待されている[6,7]。たとえば、米国では 2016 年に「がんムーンショット」計画（Cancer Moonshot Initiative）が行われ、現在 All of Us Research Program として 80 万人を超える世界的ゲノムコホートに発展している。同プログラムでは肺がん、乳がん等において類似性評価と秘匿性評価の両面から品質保証を行った検証済み合成患者データの提供が開始されている[8]。我が国の個人情報保護委員会のガイドラインに基づく一般的な加工手法として、以下が提案されている。

- ・ 項目削除／レコード削除／セル削除
- ・ 一般化（上位概念への置換、数値の丸め処理）
- ・ トップ（ボトム）コーディング
- ・ ミクロアグリゲーション
- ・ データ交換（スワップ）
- ・ ノイズ（誤差）付加
- ・ 疑似データ生成

合成患者データは、この「疑似データ生成」に当たる技術である。2024 年のガート

ナー生成 AI ハイブ・サイクルにおいても、合成患者データ生成技術は注目される分野として位置づけられており、我が国においても対応が必須であると考えられる。

加えて、米国疾病予防管理センター (CDC) は 2019 年末に DMI を開始し、COVID-19 パンデミック対策として 10 億ドル以上の緊急資金を投入した。「Data saves lives. Better data saves more lives. “データは命を救う。より良いデータは、さらに多くの命を救う”」という理念のもと、ウイルス感染状況モニタリングなどの高度な疾病監視システムやデータ分析プラットフォームが迅速に導入された[9]。また米国慢性疾患協会 (US National Association of Chronic Disease Directors) が推進する Multi-state EHR based-Network for Disease Surveillance(MENDS)は、電子カルテを用いた慢性疾患サーベイランスである。全米の電子カルテによる分散型データ収集ネットワークを通じて、2型糖尿病、喫煙、高血圧症、小児喘息、肥満等の疫学指標をリアルタイムで監視し、リスクダッシュボードや地区別有病率推定ダッシュボードを提供している[10]。

このように、医療・健康ビッグデータや AI 技術の発展を背景に、多様なデータソースを組み合わせたデータ活用と社会基盤整備が世界中で進められている。

D. 考察

生活習慣病の発症・重症化予防に関する介入プログラム開発における PHR データ活用の課題として、PHR サービスとして医療機関との情報連携にかかるコストの低減と、複数サービス間におけるポータビリティの確保が必要である。ポータビリティを容易にする標

準化を国や団体において進めることが期待される一方で、広く PHR と医療機関とのデータ連携による医療の質向上を実現するためには、従来型のデータを抱え込むモデルではなく、「本人の意思の下で 自身の健康に関するデータのやり取りが可能な社会」の未来像を社会的に共有する必要がある。社会基盤としての PHR データ流通基盤の基本機能と運用ルールを定め、医療の現場でリアリティのある実証と課題の検証を行った上で社会実装を促すために、本事業の役割は重要である。

加えて、生成系 AI を活用した合成患者データがプライバシー保護と研究有用性を両立する革新的技術として期待され、国際的に実装が進んでおり、我が国においても戦略的導入を検討すべき時期にある。また電子カルテを用いた公衆衛生監視システムの整備も劇的に進展している。これらに PHR データを加えることで、より高度な健康に関する社会基盤を構築できる可能性がある。一方で、技術的課題とともに倫理・法的枠組みの整備が不可欠であり、学際的アプローチによる継続的研究が必要である。

E. 結論

PHR アプリケーション「健康日記」を用いた介入研究の成果を迅速に反映し、システム面で研究実施のサポートを行うと共に、多様なデータソースを組み合わせた健康に関するデータ活用のための社会基盤整備に貢献したい。

【参考文献】

[1] Lee K K, et al. Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis,

and modelling study BMJ 2022; 377: e068424

[2] Digital Medicine Society (DiMe) Sensor Data Integration Project.

<https://dimesociety.org/access-resources/sensor-data-integrations/>

[3] 一般社団法人 PHR 普及推進協議会
民間事業者の PHR サービスに関わるガイド
ライン (第 2 版) . https://phr.or.jp/wp-content/uploads/2022/10/guideline_20221021.pdf

[4] Brokamp C, et al. Decentralized and Reproducible Geocoding and Characterization of Community and Environmental Exposures for Multi-Site Studies. *Journal of American Medical Informatics Association*. 2018; 25(3). 309-314.

[5] Sara Jackson. UCLA/UCSF look to build research database from health app data. 2011
<https://www.fiercehealthcare.com/mobile/ucla-ucsf-look-to-build-research-database-from-health-app-data>

[6] El Emam K, Jonker E, Arbuckle L, Malin B (2011) A Systematic Review of Re-Identification Attacks on Health Data. *PLoS ONE* 6(12): e28071.

[7] Akiya I, Ishihara T, Yamamoto K. A Comparison of Synthetic Data Generation

Techniques for Control Group Survival Data in Oncology Clinical Trials: Simulation Study. *JMIR Medical Informatics*. 08/05/2024:55118

(forthcoming/in press) DOI: 10.2196/55118

[8] Ziqi Zhang, Chao Yan, Bradley A Malin, Keeping synthetic patients on track: feedback mechanisms to mitigate performance drift in longitudinal health data simulation, *Journal of the American Medical Informatics Association*, Volume 29, Issue 11, November 2022, Pages 1890–1898,

[9] U.S. Centers for Disease Control and Prevention: CDC. Data Modernization.

<https://www.cdc.gov/data-modernization/php/about/index.html>

[10] US National Association of Chronic Disease Directors. MENDS.

<https://chronicdisease.org/cphl/technical-assistance-hub/data-modernization/mends/>

F. 研究発表

なし

G. 知的財産権の出願・登録状況

該当せず