

2022年04月28日



# 社会政策課題研究所

レポート

No.008

健康・医療情報の活用

社会政策課題研究所

所長 江崎 禎英

### 健康・医療情報の活用－ PHR データの分析可能性

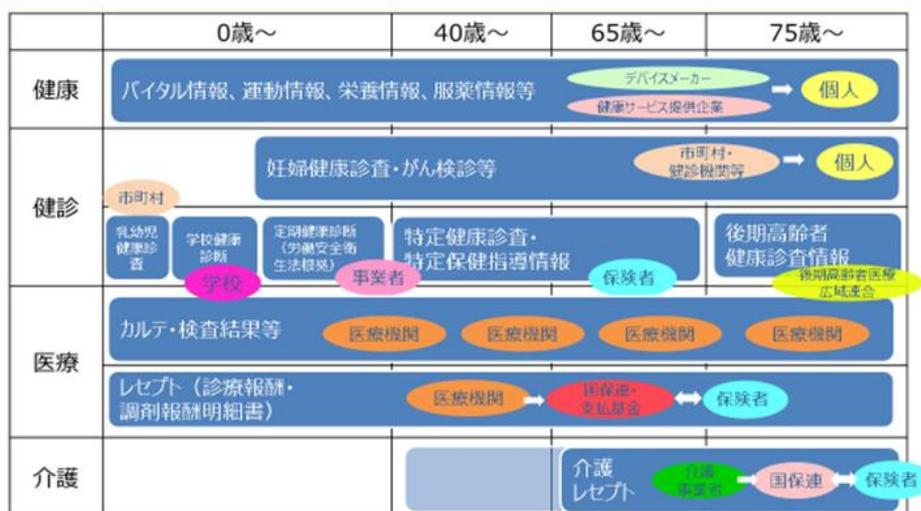
高齢化の進展とともに、如何に健康で長生きするかが重要なテーマとなるなか、個人の健康や医療に関するデータ（PHR：Personal Health Record）活用への期待が高まっています。

近年コンピュータの処理能力が飛躍的に向上した結果、これまで取り扱うことの出来なかった膨大なデータの処理が比較的短時間でできるようになりました。このことは、各分野においてそれまで専門家の経験則でしかなかった情報に一定の根拠を与えるとともに、それまで見落とされてきた事実気付くヒントを与えるものと期待されています。

しかしながら、PHR を十分に活用するには乗り越えなければならない課題も多く、本レポートでは、PHR を活用するに当たっての留意事項についてまとめます。

#### ライフステージと健康・医療・介護情報

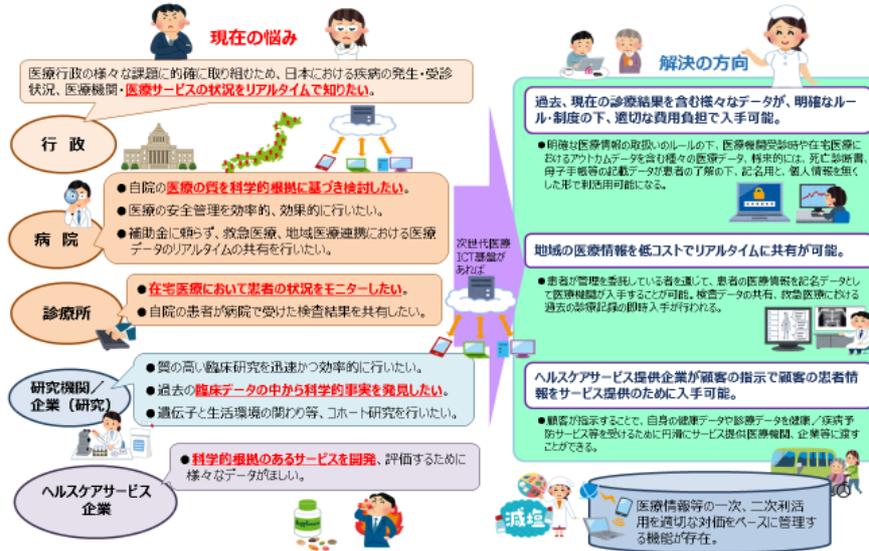
○ 健康、健診、医療、介護の各分野（縦軸）において、乳児期から高齢期までの個人の生涯における健康情報（横軸）が収集・管理・活用されている。



#### 1. PHR の現状

PHR の存在という点において、日本は世界で最も条件の整った国と言えるでしょう。乳幼児期から高齢期まで、生涯に亘ってデータが得られており、しかもその質は非常に高いのです。こうしたデータを連結し活用することによって、行政においても、病院においても、診療所においても、研究機関においても、さらにはヘルスケアなどの分野においても、様々な課題に答えが出せるとの期待が寄せられています。こうした情報を ICT によって連結すれば、次世代の医療を構築できるとの思いでこれまで様々な取り組みが行われてきました。

## 健康・医療情報の活用が目指すもの



しかしながら、現実には一部地域で独自のデータ連携によって健康管理サービスが立ち上がるなど、個別のシステムが乱立する状況になっています。

## 2. レガシーシステムの存在

なぜそのようなことになっているのでしょうか。日本における情報化は、従来の業務体系をそのままITに置き換えてしまうケースが一般的であったため、これがいわゆる「レガシーシステム」となり、データ連携やシステム統合の障害になっています。情報共有の基盤となる汎用システムであるべきところを、利用者の要望に合わせてカスタマイズし過ぎた結果、複数の医療機関どころか医療機関内部でもデータが繋がらない状態になっていることが珍しくありません。

因みに、システムを繋ぐという点だけで言えば、構造が違うデータの上位に共通のタグをつけることで連結したように見えます。しかしながら、実際の研究活動に使用する段階では、それぞれのデータベースに戻って個別のデータを比較検証するという作業をやり直す必要が生じてしまいます。システム上でデータを比較検証するには、基本的なデータ構造まで戻って共通化する必要があるのです。特にMRIやCTも含めて、検査機器毎にデータ構造が違うという点を理解しておかないと、闇雲にデータを集めただけでは実際に有意な分析を行うことは難しいのです。

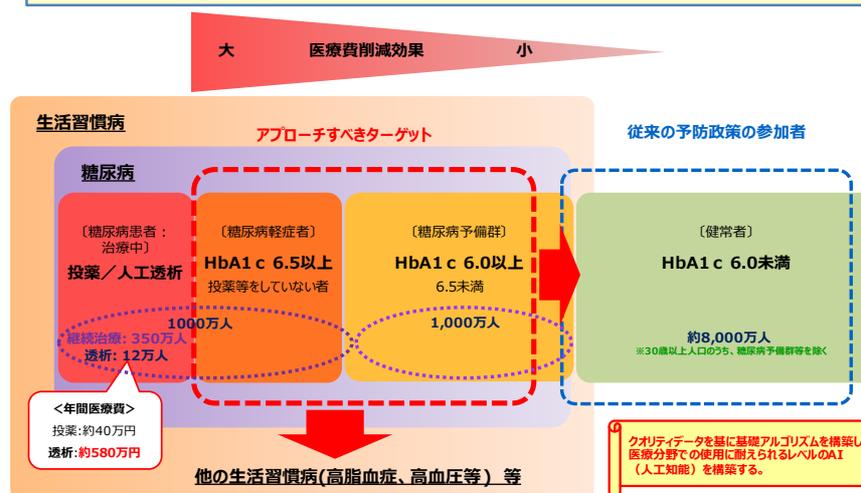
## 3. データの偏り

現在、ウェアラブル機器などを用いた健康プログラムが盛んに開発されていますが、比較的大量に集めることが出来るPHRの多くは、健康に関心のある比較的健康な方のデータに偏っています。こうした情報を用いて生活習慣病予防のためのプログラムを開発しようと

しても、本来予防に取り組まなければならない人々のデータが含まれていないという課題に直面します。そうしたデータをいくら分析しても、糖尿病重症化予備群に適したプログラムを作成することはできません。健康な方々のデータをもとに運動プログラムを作り、糖尿病患者にむやみに運動させることは適当ではないのです。

### 健康・医療情報の活用 – ビックデータからクオリティデータへ –

- 本人同意を前提に、IoTによる取得されるデータの連携、蓄積基盤を構築(交換規約やデータ交換様式の検討、試行)。
- 医学的に確立された糖尿病診断指標(HbA1c)を用い、治療等に用いられる手法(対照実験等)により効果を検証



#### 4. データの項目数

かつて糖尿病の重症化予防プロジェクトを行った際には、研究者側からは 160 を超える項目のデータ取得が求められました。しかしながら、特殊な環境や検査機器が無ければ取得できないデータも多く、日常生活において長期継続的に、かつ、正確にデータを収集できるとの観点から検証を重ねた結果、最終的に 4 項目に絞られました。データ項目を増やしても、測定方法の違いや測定誤差が積み重なってしまうことで、却って有意な分析の妨げになるのです。PHR の活用について検討する場合、研究に必要なデータと安定的に取得できるデータのギャップについて十分な理解が必要です。

#### 5. データの粒度

データの比較分析には、データフォーマットの標準化が重要と言われてきました。異なるシステムで入力されたデータは、接続に必要な本人確認や単位の調整、測定誤差の調整といったいわゆる「データクレンジング」に莫大なコストが掛かるのです。これに加え、IT システムの能力向上は新たな問題を発生させています。

先に述べた糖尿病の重症化予防プロジェクトでは、8つのチームが全く同じデータフォーマットで 3 カ月間にわたって体重、血圧、活動量といった生活習慣データを収集したのですが、結果的にこれらのデータは全く接続できませんでした。

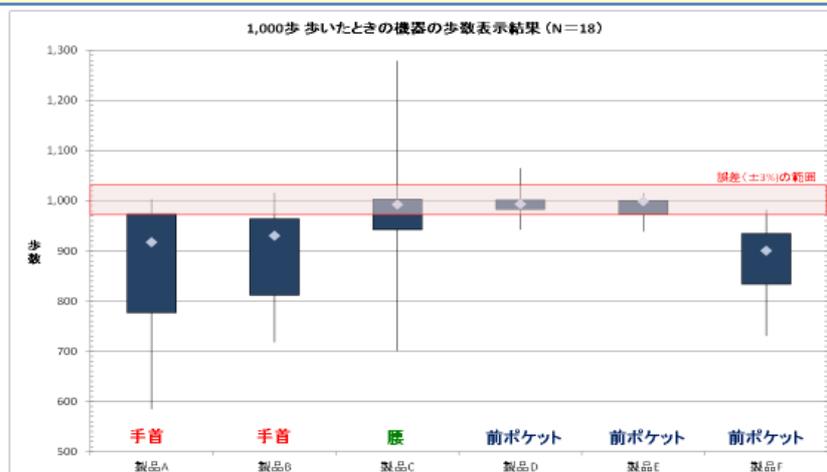
## 交換規約で取り扱うデータ項目詳細（一部抜粋）

項目名	物理名(英字)	型	桁	必須	備考
計測日時	observation_time	日付	14	●	yyyyMddHhmmss 計測データの計測した日時 ※秒、分、時を計測していない場合は該当部分に 00を入力
歩数	step	数値	5	●	単位: 歩、整数
歩行時間	walk_time	数値	4	▲	単位: 分、整数 ※計測している場合のみ入力
中強度時間	moderate_intensity_activity	数値	4	▲	単位: 分、整数 ※計測している場合のみ入力
消費カロリー	calories	数値	6	▲	単位: kcal, XXXX.X ※計測している場合のみ入力
総消費カロリー	total_calories	数値	6	▲	単位: kcal, XXXX.X ※計測している場合のみ入力
機器メーカーコード	model_manufacturer_code	文字列	50	●	メーカーコード、機器から出力される通信仕様で決めた番号等 ※機器から取得できない場合は事務局が作成する機器リストの機器メーカーコードを入力
機器型番	model_number	文字列	50	●	機器から出力される型番コード ※機器から取得できない場合は事務局が作成する機器リストの機器型番コードを入力、スマートフォンアプリの場合は、アプリケーション名とバージョンを入力 例: HealthKit X.X
機器識別ID	production_specification_serial	文字列	50	●	機器識別番号、製造番号などの機器を特定できる番号 ※機器から取得できない場合は台帳で管理している機器管理番号で代替。その場合には、先頭に「コンソーシアムID」を付けて、「コンソーシアム ID」+「機器管理番号」とする、スマートフォンアプリの場合はスマートフォンを特定できる識別番号(例: Macアドレス)
活動量集計時間	total_active_duration	数値	4	●	活動量の記録集計時間(分)を記載 ※活動量計の場合必須
データ入力コード	data_input_code	文字列	1	●	0: システム入力(機器入力) データ、1: 手入力(アプリケーション画面入力) データ、2: 機器入力データをアプリケーション側により手動で修正
活動量計装着箇所コード	measurement_position_code	文字列	2	▲	10: ポケット内に収納(胸部)、11: ポケット内(腰部)、20: 吊り下げて利用(首まわり)、21: 吊り下げて利用(ベルトまわり)、30: 身につけて利用(利き手手首)、31: 身につけて利用(利き手以外の手首)、40: 携行して利用(かばんの中に収納)、50: その他
時刻補正コード	origin_of_time_code	文字列	1	●	0: 機器の時刻をそのまま使用、1: アプリケーションにより機器の時刻を補正
更新日時	update_time	日付	17	●	yyyyMddHhmmssSSS アプリケーション内(データベース)でのデータ登録更新日時、データ更新時の過去データとの整合性チェックに利用する。初回であれば登録日時となり、次に同じ計測日時についてデータ修正のため再送する場合は、各コンソに保持しているデータの更新日時を設定する。保持していない場合には、送信日時と同様となる。

主な原因は、ウェアラブルなどの活動量計から送られてくるデータの送信頻度(データ粒度)のバラツキです。例えば A のウェアラブルは 1 秒ごとにデータが送信され、B のウェアラブルでは 5 分に 1 回、C のウェアラブルでは最初の 30 分を除いて 15 分に 1 回送信されるため、クラウドに自動的に蓄積されたデータを相互に比較分析することが出来なかったのです。既に存在するウェアラブルなどの測定機器は、こうしたデータの粒度が整えられていません。同じ会社のものでも機種によって異なります。たった 3 項目のデータですらこの状態ですから、様々な機器で収集されたデータの比較検証が極めて困難であることは容易に想像できると思います。

## 測定データの信頼性について

- 全製品とも平均誤差率は実測値を下回る結果であり、**-0.6%~-13.3%**であった。
- **製品D、製品E**の測定誤差は第1、第3四分位点が**JIS基準範囲内**であった。
- 製品A、製品Cは**参加者による誤差が大きく**、その差は400-600歩程度あった。
- 製品B、製品Fは測定値の多くが**800歩-900歩**となった。



## 6. ビッグデータの落とし穴（ストレージ能力の向上に伴う課題）

ビッグデータという言葉がもてはやされ、沢山のデータがあれば「色々なことが分かる」と期待されました。特に、情報技術の急速な発展に伴ってデータを保存するストレージの能力が飛躍的に向上しました。この結果、様々なセンサー機能の向上とも相俟って、一人の患者に関する膨大なデータが簡単に蓄積されるため、却って分析が難しくなるという悪循環に陥っています。IT システムの能力が向上するほど、データの蓄積が加速度的に増えてしまい、研究活動に使い憎くなる状況で膨大なリソースが浪費されるという皮肉な結果になりつつあるのです。

## 7. 今後のシステムに向けて

PHR を始めとする健康・医療情報への期待が益々高まるなか、安全であることの意味、情報を活用するための制度の在り方について、その必要性や課題について十分に理解しておくことが大切です。

特に、今後疾患の中心が生活習慣病と老化になるのであれば、それほど沢山の種類や機微な情報も必要ないかもしれません。ただ、ひょっとしたら違う病気を見落としているかもしれないときのための参考情報はどこかに保管しておけばよいのです。念のための情報は結局見ないままになることも多いため、できるだけ安いコストで保管することが肝要です。そのように考えて医療情報システムを見直せば、健康・医療分野における PHR の活用性が広がるとともに将来的な AI 活用の可能性も現実味を帯びてくるのではないのでしょうか。

今後の健康・医療情報システムは、どうあるべきかを考えた場合、できるだけ医療関係者や利用者本人が簡単にアクセスできることを心がけるべきです。迅速性と正確性が確保でき、簡便な入力方法にした上で、データの共通化は必須です。入力段階での統一を図るとともに、あとで使うことを意識しながらデータの蓄積を考え、匿名化はしないことです。セキュリティ確保のための暗号化も、医療分野では解読ミスが致命傷になるため極力避けるべきでしょう。