

令和2年度厚生労働科学研究費補助金  
(政策科学総合研究事業(統計情報総合研究事業))  
分担研究報告書

高齢者ヘルス領域における海外データリンケージの状況

研究代表者 伊藤 智子 筑波大学医学医療系 助教  
研究分担者 田宮 菜奈子 筑波大学医学医療系 教授  
研究分担者 森 隆浩 筑波大学医学医療系 准教授

研究要旨

高齢者ヘルス領域におけるリンケージデータを用いた研究の実績について、PubMed を用いて、検索語「Elderly Care Linkage」によって文献を検索した。検索にヒットした文献のうち、主要なリンケージデータソースを抽出し、そのデータ提供の方法や研究実績についてレビューした。その結果、7件の大規模データベースについてレビューし、そのうち3件が米国における事例であり、英国1件、EU1件、オーストラリア1件、カナダ1件であった。これら海外のデータベース運営方法は、我が国のように各官庁の既存の組織内の限られたソースの中で管理されていることに比較して、多くの案件を迅速に処理することを可能にしていると考えられた。

A. 研究目的

海外のデータリンケージの実態を把握すべく、今回は、特に高齢者ヘルス領域における海外の著名な大規模データリンケージのおよびそのデータを用いた研究実績についてレビューを行った。

B. 研究方法

高齢者ヘルス領域におけるリンケージデータを用いた研究の実績について、PubMed を用いて、検索語「Elderly Care Linkage」によって文献を検索した。検索にヒットした文献のうち、主要なリンケージデータソースを抽出し、そのデータ提供の方法や研究実績についてレビューした。

C. 研究結果

検索の結果、254件(2021年2月時点)が該当し、そのうち、リンケージデータに関する文献は96件であった。SEER-Medicare linked database(米国)、Clinical Practice Research Datalink(英国)、Survey of Health, Ageing and Retirement in Europe(EU)、The Centre for Health Record Linkage(オーストラリア、ニューサウスウェールズ州)、ICES(カナダ、オンタリオ州)、Rochester Epidemiology Project medical record(米国、ミネソタ州オルムステッド郡)の7種類のデータベースが見いだされ、各データベースにおける運営体制やデータ内容、データ利用方法などを比較した。報告事例の

7件中3件が米国における事例であり、英国1件、EU1件、オーストラリア1件、カナダ1件である。データ間のリンケージ方法は社会保障番号などの個人番号を用いた **Deterministic Linkage** が4件であった。また、データの利用率については、4件が有料であった。データベース提供事業の運営組織は様々であり、国立研究所や大学内の特別部局、あるいはNPO組織などによるものであった。

以下に各データベースの詳細を説明する。

## 1. 全国、国際規模の医療突合データベース

### 1) SEER-Medicare database (米国)

米国の **SEER-Medicare database** は、**Surveillance, Epidemiology and End Results (SEER)** と **Medicare** の利用情報を統合した、高齢がん患者に関する全国規模の医療データベースである。<sup>1 2</sup>

**SEER-Medicare database** は、高齢がん患者に関する研究支援を目的として、**SEER** を管理する政府機関である **National Cancer Institute (NCI)** と、**Medicare** の請求データを保有する **Centers for Medicare & Medicaid Services (CMS)** によって、1991年に設立された。現在もこの2機関が管理および突合を行う。1991年の最初の突合から、2018年までに8回のフォローアップが行われている。

米国ではがんが新規診断された場合、医療提供者は所属地域の **Cancer registry** への報告が求められている。NCIが執行する **SEER** プログラムは全国の **Cancer**

**registry** の情報を収集する。がんの新規診断情報に加えて、人口統計的情報（診断時の年齢、性別、人種、民族的帰属など）、診断された日付、腫瘍の特徴、生存情報、死因に関するデータを内包する。2018年以降、**SEER** は米国人口の約35%が居住する21州の **Cancer registry** から情報を収集している。米国全体と比較すると、**SEER** 登録者の人口構成は人種的、民族的多様性に富み、経済的に恵まれない人々を含む傾向にある。

**Medicare** は、米国の65歳以上人口の約94%を対象とした健康保険プログラムである。出資は **CMS** が行う。**SEER-Medicare database** には、**Medicare** が記録する病院、外来治療施設、医療従事者、ホスピス、在宅医療サービスの請求情報が含まれる。

補助的データとして、**SEER** と **Medicare** に登録された患者の郵便番号で国勢調査局の情報を参照することで、社会経済学的情報（地域の世帯収入の中央値、教育レベル）を導入している。また、患者に対応した医療提供者に関するデータベースや、患者の合併症インデックスも含まれる。

**SEER** と **Medicare database** における個人の突合には、各データに含まれた社会保障番号 (SSN)、氏名、性別、生年月日、死亡日を使用した **Deterministic Linkage** が用いられる。2018年の最新版時点で、**SEER** 登録者の96%が **Medicare** のデータとリンケージされている。

研究例としては、**SEER** のがん診断情報と **Medicare** の人口統計データを組み合わせることで、がんの初診断時のステージの

傾向と、人種およびスクリーニングサービスの利用頻度の関連性を検証したものである。<sup>3</sup>

## 2) Health and Retirement Study (HRS) (米国)

Health and Retirement Study (HRS) は、米国に住む 51 歳以上の 37,000 人以上を対象とした長期パネル調査である。<sup>4</sup> 高齢化に伴って変化する健康と経済状況に関するデータの提供を目的として、国立衛生研究所傘下のアメリカ国立老化研究所 (National Institute on Aging, NIA) によって 1992 年に創立された。HRS は高齢者の健康と経済状況を対象とした米国初の長期的研究であり、参加者の収入と豊かさ、健康や認知機能と医療サービスの利用、職業と引退、家族とのつながりといった広範なテーマに着目する。

長年にわたり、米国の科学者と政策立案者は米国民の急速な高齢化を予測していた。1980 年代、NIA は高齢化研究をサポートする情報源の必要性を認識し、Dr Richard Suzman の指揮のもと、HRS の基本的枠組みが確立された。1990 年に米国議会が NIA に HRS の創設を指示した。HRS は NIA の社会行動研究部門 (Division of Behavioral and Social Research, BSR) と、社会保障局 (Social Security Administration) の共同出資によって運営される。

HRS の調査は、BSR と共同契約を結ぶミシガン大学の Institute for Social Research (ISR) によって行われ、1992 年から 2016 年にピックアップされた世代別の 6 つのコホートをフォローアップする。

<sup>5</sup>参加者の半数は対面のインタビュー、体力測定、バイオマーカー測定によって情報を取得し、残りの半数は電話によるコア・インタビューのみを行う。フォローアップ調査は 2 年毎に行われる。HRS は National Death Index (NDI) との突合によって参加者の生存情報を確認し、死亡時には配偶者や子供にインタビューを行い、死亡した参加者の医療費、家族の交流、資産の処分などについて調査する。

HRS の個人データは社会保障、Medicare、復員軍人援護局 (Veteran's Administration)、雇用年金、National Death Index の死亡情報、アメリカ合衆国退役軍人省 (Veterans Affairs) の医療記録といった行政データとの突合が行われる。行政データとの突合には参加者の同意が必要であり、Medicare との突合の同意率は 78% から 84% である。また、参加者の年金情報と雇用主の履歴を参照できる Employer Identification Number (EIN) の取得が試みられている。

HRS と行政データの突合のプロセスは対象となるデータに応じて異なる。Centers for Medicare & Medicaid Services (CMS) のデータとの突合<sup>6</sup>では、第三者機関である Acumen に HRS と CMS がデータを送信し、Acumen が個人情報の突合を行ってファイルを生成する。退役軍人省のデータの突合<sup>7</sup>では、HRS のファイルが VA Information Resource Center (VIReC) に送信され、VIReC によって個人情報の Probabilistic linkage が行われる。

HRS のデータを用いた突合研究例としては、HRS が取得したアルコールの消費

量と NDI の死亡データの突合データを使い、アルコール摂取と高齢者の死亡率の相関性を調査した研究<sup>8</sup>や、HRS の参加者情報、NDI の死亡データ、Medicare の請求データを突合し、参加者の最後の 6 ヶ月の医療支出における人種的、民族的傾向の差異を調査した研究<sup>9</sup>がある。

### 3) Clinical Practice Research Datalink (CPRD) (英国)

Clinical Practice Research Datalink (CPRD) は、英国内で最も大規模な研究用医療データベースである。英国の保健省 (Department of Health and Social Care) の下部機関である医薬品・医療製品規制庁 (Medicines and Healthcare products Regulatory Agency, MHRA) と国立衛生研究所 (National Institute for Health Research) の助成を受けている。<sup>1011</sup>

CPRD は、デジタル健康記録データを用いた医療研究における英国独自のポジションを構築するという政府の戦略の一環として、2012 年 4 月に設立された。<sup>12</sup>

英国に在住記録のある累計 5000 万人 (現在の居住者では 1500 万人) の医療関連データを、全国の一般診療医から収集する。<sup>13</sup>データの収集・管理は MHRA が行う。<sup>14</sup>

内包するデータは、外来診療、入院、救急医療、治療アウトカム、死亡、がん診断、メンタルヘルスなどの医療情報のほか、郵便番号を基軸とした居住地域の貧困度などの人口動態学的情報を含んでいる。

CPRD における個人情報突合は、保健省の下部機関である NHS Digital が行う。<sup>15</sup>国民保険番号、郵便番号、生年月日、

性別を用いた 8 段階の Deterministic Linkage によって個人が突合される。突合後、個人を特定する情報は消去される。

CPRD の情報を取得するためには、プロトコル申請書の提出、Independent Scientific Advisory Committee (ISAC) による承認、およびライセンス支払いが必要である。<sup>16</sup>

利用ライセンスは複数の研究に使うことができる 12 ヶ月間の定額制と、単独研究用のライセンスに分かれており、アクセス可能なデータベースと利用者の性質に応じて種別と料金が異なる。また個人の突合には追加料金を請求する。<sup>17</sup>

CPRD を用いた研究例としては、糖尿病患者の低血糖症による入院傾向について、高齢患者とそれ以外の年齢層の患者の比較研究が行われた。<sup>18</sup>若年患者と比較して、高齢の糖尿病患者の低血糖による入院率は 1998 年から 2013 年に上昇していることが確認された。

### 4) Survey of Health, Ageing and Retirement in Europe (SHARE) (EU)

Survey of Health, Ageing and Retirement in Europe (SHARE)<sup>19</sup>は、EU の大半の所属国およびイスラエルの 50 歳以上の人口を対象とした、健康、社会経済学、および社会と家族ネットワークという複数の視点における長期的研究である。

SHARE は、欧州連合内の高齢化人口に関する科学的エビデンス取得への強い関心を背景に、複数の領域の科学者による長期的サーベイのインフラとして 2002 年に発足された。2011 年から SHARE は初の European Research Infrastructure

Consortium (ERIC)となり、重要な国際組織としての優位性を得た。調査参加者は約14万人であり、2004年から2017年までに7回の調査が行われている。<sup>20</sup>SHAREのデータの統合はマックス・プランク社会法・社会政策研究所 (Max Planck Institute for Social Law and Social Policy) の1部門である、Munich Center for the Economics of Aging (MEA)が行う。

2122

SHAREの調査は質問票、体力測定、バイオマーカーの測定によって行われる。質問票では参加者の年齢、性別、既往歴、入院歴、ワクチン接種歴のほか、結婚状態、雇用や経済状態を調査し、身体測定では握力、心肺機能、歩行能力などを記録する。身長、ウエスト、血圧や Dried blood spots によるバイオマーカー測定はドイツの対象者に限定される。参加者が死亡した場合は、代理人との対面あるいは電話によるフォローアップで参加者の最終生存年の情報を取得する。

SHAREが収集するデータは幅広い領域に渡るが、外部の行政データとの突合を行うことによって、より正確かつ完全な情報を取得できる。SHAREはドイツ年金基金 (German Pension Fund, DRV) と協力しており、DRVが保有するサーベイデータとの突合が可能である (SHARE-RV)。SHARE-RVにおける個人の突合には、Social Security Number を使用した Deterministic Linkage を採用する。<sup>23</sup>

デンマーク (REGLINK-SHAREDK)、オランダ (Linkage SHARE NL) の行政データとの突合も行われている<sup>24</sup>が、SHARE-RVも含めて、利用されている研

究例は少ない (2021年5月現在のPubMed検索では"SHARE-RV"のヒットが2件、REGLINK-SHAREDK、Linkage SHARE NLはヒットせず)。その理由としては、突合データの設立が新しいことや、SHARE自体のデータが広範にわたるため、SHAREのみで研究が完結する傾向が強いことが考えられる。

行政データとSHAREの突合研究の例として、ドイツにおける50歳から65歳の労働者のストレスと引退年齢を調査したもの<sup>25</sup>がある。この研究ではSHARE-RVを用いているが、データの突合には参加者の同意が必要であり、サンプル数が限定されてしまうことが論文内で言及されている。このような背景から、ある1国のデータを調査する場合、EUの横断研究であるSHAREよりも、他の国内の調査を用いた方が多くの正確なサンプルが確保できることが、SHAREの突合研究が少ない理由の1つとして考えられる。

SHAREを使用した高齢者に関する研究例としては、65歳以上の対象者の多剤併用の傾向と関連因子についての17カ国の横断的研究がある。<sup>26</sup>5種類以上の薬剤を併用する患者の割合は、居住国によって26.3%から39.9%まで分かれていた。この研究のように、複数の参加国の状況を同じ尺度から横断的に調査できることがSHAREの長所として認められる。

2. 地域・州規模の医療突合データベース  
1) The Centre for Health Record Linkage (CHeReL) (オーストラリア)

The Centre for Health Record Linkage (CHeReL)<sup>27</sup>は、オーストラリアのニュ

ーサウスウェールズ州およびオーストラリア首都特別地域の約6900万人を対象とした医療データベースの突合ユニットである。

CheReLは研究者・ヘルスプランナー・政策立案者に医療に関するリンケージ情報を提供し、健康に関連する活動を支援することを目的として、ニューサウスウェールズ州保健省（The NSW Ministry of Health）によって2006年に設立された。<sup>2829</sup>同省が管理および資金提供を行う。<sup>3031</sup>

CHeReLのデータユニットが医療機関から取得する医療データは、緊急医療、入院、出生、死因、がん診断などである。<sup>32</sup>

各データの個人の突合は、CHeReLのデータリンケージユニットによって施行される。<sup>33</sup>研究者からのデータ提供の要望があった際は、まず対象となる医療データから個人情報と分離し、氏名、性別、生年月日などの個人情報をもとに独自の個人識別番号を適用する。その後、個人識別番号で各個人の医療データを突合し、研究者に提供するというプロセスを経る。この突合は個人情報をベースとしており、Choicemakerというソフトウェアを使用したProbabilistic linkageである。<sup>3435</sup>CHeReLは突合の偽陽性率0.5%を目標として掲げている。

CheReLのデータへのアクセスには、研究プロトコルと申請書の承認が必要である。<sup>36</sup>料金はプロジェクトの性質と利用データ量に依存する。<sup>37</sup>公式サイトが紹介する例では、2種類のデータベースからの50,000人分の突合情報の取得には約5,000豪ドルが請求される。また、外部のデータベースとの突合には追加料金が求

められる。

高齢者を対象とした研究例として、2007年にニューサウスウェールズ州で死亡した70歳以上の34,556名を対象に、入院記録、がん登録情報、死亡記録および死因情報を突合して、最終生存年における医療サービス利用の特徴が調査された。<sup>38</sup>対象者の82%が最終生存年に1回以上入院しており、21%は4回以上入院していたことが判明した。

## 2) ICES（カナダ）

ICESは、カナダ・オンタリオ州の医療関連データを収集し、医療に関する意思決定のエビデンスとして提供することを目的として、1992年に設立された非営利の研究機関である。<sup>39</sup>ICESはオンタリオ州のMinistry of Health and Long-Term Careによる公的出資を受けており、民間部門からは出資されていない。<sup>40</sup>

ICESはオンタリオ州の公的出資を受けた健康管理機関からデータを収集し、2013年時点で約1300万人分の医療記録を包有する。<sup>41</sup>がん、脳卒中、心臓病といった特定疾患のデータベースのほか、慢性疾患患者のコホート、入院、手術、リンケージ可能なCanadian Community Health Surveyのような大規模な健康調査データ、国勢調査、人口動態学的データを含む。<sup>42</sup>

ICESにおける個人データの突合は、患者個人を特定する情報を排除して行われる。ICESは協力機関から医療データを取得後、氏名などの個人情報を消去し、代わりに独自の識別番号（ICES Number）を割り振った後に他のデータとの突合を行

う。<sup>43</sup>

高齢者を対象とした研究例としては、2010年から2016年の間にステージⅠの乳がんの診断を受けた65歳以上の患者を対象に、内分泌療法、放射線療法、補助的治療のアウトカムの差異が調査された。<sup>44</sup> 局所再発リスクは補助的治療を受けていないグループで有意に高かった。また、内分泌療法と放射線療法のリスク差は小さかった。この研究ではがん患者、入退院、日帰り手術、人口動態学情報のデータベースを突合している。

### 3) Rochester Epidemiology Project medical records - linkage system (REP) (米国)

Rochester Epidemiology Project medical records - linkage system (REP) は、ミネソタ州オルムステッド郡の人口をカバーする医療情報突合プログラムである。<sup>45</sup> 明確に定義された人口における長期的な医療データの提供を目的として1966年に設立された。設立から47年間はアメリカ国立衛生研究所 (National Institutes of Health, NIH) が、現在はアメリカ国立老化研究所 (National Institute on Aging, NIA) が資金を提供している。2012年時点で50以上の医療提供機関が参加し、502,820名の医療データを格納する。

20世紀前半から、メイヨ・クリニック (Mayo Clinic) は患者番号を用いた独自の医療記録を保有していた。この医療記録を使用した研究に影響を受けた神経学と伝染病学の専門家である Leonard T. Kurland は、1966年にNIHから資金調達を行い、ロチェスター市とオルムステッ

ド州の他の医療施設に診断コードのインデックスを創設したことで、メイヨ・クリニックの医療記録との突合が可能になった。これがREPの創設である。<sup>46</sup>

REPに参加する全ての医療提供者は、患者の人口統計学的な情報 (氏名、性別、生年月日、住所)、Mayo Clinic patient numberなどの医療提供者固有の患者個人番号、および医学的状态や外科処置についての診断コードをREPに提供する。年4回もしくは年2回の周期でデータの取得を行う。診断・外科処置コードが対応する患者データと適切に突合された後に、REPのインデックスに記録される。

死亡日については、医療提供者からは他の患者情報の習得時に、ミネソタ州の死亡登録簿からは年4回の周期で取得を行う。死因情報はミネソタ州が提供する。また、オルムステッド郡における検死情報も提供される。

REPのデータを研究利用するためには、Mayo Clinic、Olmsted Medical Center (OMC)、Institutional Review Boards (IRBs)による研究プロトコルの承認が必要である。<sup>47</sup> 承認を受けた研究者は、REP Browser という専用のアプリケーションを使用して、個人単位に突合された患者情報にアクセスすることができる。

複数の医療提供期間の患者情報の突合方法として、REPは1966年の設立時から2002年まで、研究ごとに特定の患者情報をマッピングする Probabilistic Linkageを採用していたが、煩雑でコストが掛かり、エラーの発生率が高かった。2002年以降は、SAS software package (SAS Institute, Inc., Cary, North

Carolina) の Soundex phonetic filing system を採用することで、誤記が起きやすい名前を自動的に突合することが可能になった。現在では氏名、性別、生年月日、社会保険番号といった個人情報を合計 20 段階の方法で突合し、更に REP のテクニカルチームによるチェックを加えることで、正確性を確保している。<sup>48</sup>

REP には限界も指摘されている。サンプル人口が限定的であることから、膵臓がん、卵巣がんなどの希少な疾病の研究は困難であるほか、定期的な医療記録には残されない軽度認知障害や胃食道逆流といった症状の研究も難しいとされている。また、REP の対象人口の社会経済学的な特徴は米国北中西部を反映したものであり、他地域の人種・民族的特徴とは異なることに留意する必要がある。

REP を利用した研究として、オルムステッド郡の 65 歳以上の高齢者を対象に、抗うつ剤の過剰処方に関わる予測因子が調査された。<sup>49</sup>その結果、介護老人ホームの居住、合併症、緊急外来の使用、併用薬の処方、対面以外の方法による処方などが抗うつ剤の過剰処方の関連因子として判明した。

#### D. 考察

世界有数のデータベースのリンケージ事例から、本邦のリンケージデータ活用における課題を考察した。それぞれのデータベース管理およびリンケージ処理は、半公的あるいは第三セクター的な位置付けの専門組織によって行われていた。こうした運営方法は、我が国のように各官庁の既存の組織内の限られたソースの中で管理さ

れていることに比較して、多くの案件を迅速に処理することを可能にしているとみられた。その結果、提供されたリンケージデータを用いてタイムリーな研究結果に繋がり、データ活用の有効なフローの構築に貢献していると考えられた。

#### E. 結論

高齢者ヘルス領域におけるリンケージデータを用いた研究の実績について、PubMed を用いて、検索語「Elderly Care Linkage」によって文献を検索した。検索にヒットした文献のうち、主要なリンケージデータソースを抽出し、そのデータ提供の方法や研究実績についてレビューした。その結果、7 件の大規模データベースについてレビューし、そのうち 3 件が米国における事例であり、英国 1 件、EU1 件、オーストラリア 1 件、カナダ 1 件であった。これら海外のデータベース運営方法は、我が国のように各官庁の既存の組織内の限られたソースの中で管理されていることに比較して、多くの案件を迅速に処理することを可能にしていると考えられた。

#### F. 研究発表

1. 論文発表 なし
2. 学会発表 なし

#### G. 知的財産権の出願・登録状況（予定を含む）

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

---

参考文献

<sup>1</sup> Brief Description of SEER-Medicare Database.

<https://healthcaredelivery.cancer.gov/seermedicare/overview/>, Accessed on May 11 in 2021

<sup>2</sup> Lindsey Enewold, et al. Updated Overview of the SEER-Medicare Data: Enhanced Content and Applications. *J Natl Cancer Inst Monogr.* 2020 May 1;2020(55):3-13.

<sup>3</sup> Marian E Gornick, et al. Associations of race, education, and patterns of preventive service use with stage of cancer at time of diagnosis. *Health Serv Res.* 2004 Oct;39(5):1403-27.

<sup>4</sup> Amanda Sonnega, et al. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol.* 2014 Apr;43(2):576-85.

<sup>5</sup> Getting Started: How to Use This Site | Health and Retirement Study, <https://hrs.isr.umich.edu/about/how-to-use-this-site>, Accessed on May 11 in 2021

<sup>6</sup> HRS Medicaid Analytic Extracts (MAX) and Summary Files Cross-Reference | Health and Retirement Study, <https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9678>, Accessed on May 11 in 2021

<sup>7</sup> Veterans Affairs (VA) Health Care

Information | Health and Retirement Study <https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9679>, Accessed on May 11 in 2021

<sup>8</sup> Katherine M Keyes, et al. Alcohol Consumption in Later Life and Mortality in the United States: Results from 9 Waves of the Health and Retirement Study. *Alcohol Clin Exp Res.* 2019 Aug;43(8):1734-1746.

<sup>9</sup> *J Health Care Poor Underserved.* 2013 Nov;24(4):1666-75. An examination of end-of-life care in a safety net hospital system: a decade in review. *J Health Care Poor Underserved.* 2013 Nov;24(4):1666-75.

<sup>10</sup> Clinical Practice Research Datalink | CPRD. <https://cprd.com/home>, Accessed on May 11 in 2021

<sup>11</sup> Emily Herrett, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015 Jun; 44(3): 827-836.

<sup>12</sup> Launch of the Clinical Practice Research Datalink - GOV.UK. <https://www.gov.uk/government/news/launch-of-the-clinical-practice-research-datalink>, Accessed on May 11 in 2021

<sup>13</sup> CPRD linked data | CPRD. <https://cprd.com/linked-data>, Accessed on May 11 in 2021

<sup>14</sup> Medicines and Healthcare products

---

Regulatory Agency privacy notice. <https://www.gov.uk/government/publications/mhra-privacy-notice/mhra-privacy-notice>, Accessed on May 11 in 2021

<sup>15</sup> How can the Clinical Practice Research Datalink (CPRD) link primary care data from multiple GP software systems without duplication? | SAPC, <https://sapc.ac.uk/conference/2018/abstract/how-can-clinical-practice-research-datalink-cprd-link-primary-care-data>, Accessed on May 11 in 2021

<sup>16</sup> Research applications | CPRD. <https://cprd.com/research-applications>, Accessed on May 11 in 2021

<sup>17</sup> Pricing | CPRD. <https://www.cprd.com/pricing>, Accessed on May 11 in 2021

<sup>18</sup> Victor W. Zhong, et al. Incidence and Trends in Hypoglycemia Hospitalization in Adults With Type 1 and Type 2 Diabetes in England, 1998–2013: A Retrospective Cohort Study, *Diabetes Care*. 2017 Dec;40(12):1651-1660.

<sup>19</sup> Axel Börsch-Supan, et al. Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol*. 2013 Aug; 42(4): 992–1001.

<sup>20</sup> The Survey of Health, Ageing and Retirement in Europe (SHARE): Waves Overview. <http://www.share-project.org/data-documentation/waves-overview.html>, Accessed on May 11 in 2021

<sup>21</sup> SHARE - Survey of Health, Ageing and Retirement in Europe - Coordination. <http://www.share-project.org/organisation/coordination.html>, Accessed on May 11 in 2021

<sup>22</sup> Social Policy (MEA) - Munich Center for the Economics of Aging. <https://www.mpisoc.mpg.de/en/social-policy-mea/>, Accessed on May 11 in 2021

<sup>23</sup> SHARE Wave 4 Innovations & Methodology. [http://www.share-project.org/fileadmin/pdf\\_documentation/Method\\_FRB\\_FINAL.pdf](http://www.share-project.org/fileadmin/pdf_documentation/Method_FRB_FINAL.pdf), Accessed on May 11 in 2021

<sup>24</sup> The Survey of Health, Ageing and Retirement in Europe (SHARE): Record Linkage Project. <http://www.share-project.org/special-data-sets/record-linkage-project.html>, Accessed on May 11 in 2021

<sup>25</sup> Jana Mäcken. Work stress among older employees in Germany: Effects on health and retirement age. *PLoS One*. 2019 Feb 4;14(2):e0211487.

<sup>26</sup> Luís Midão, et al. Polypharmacy prevalence among older adults based on the survey of health, ageing and retirement in Europe, *Arch Gerontol Geriatr*. Sep-Oct 2018;78:213-220.

<sup>27</sup> About Us – CHeReL. <https://www.cherel.org.au/about-us>, Accessed on May 11 in 2021

<sup>28</sup> Structure and Governance – CHeReL. <https://www.cherel.org.au/structure-and-governance>, Accessed on May 11 in 2021

<sup>29</sup> Mission Statement – CHeReL. <https://www.cherel.org.au/mission->

---

statement, Accessed on May 11 in 2021

<sup>30</sup> Funding – CHeReL.

<https://www.cherel.org.au/funding>, Accessed on May 11 in 2021

<sup>31</sup> Structure and Governance - CHeReL,

<https://www.cherel.org.au/structure-and-governance>, Accessed on May 11 in 2021

<sup>32</sup> Data dictionaries – CHeReL.

<https://www.cherel.org.au/data-dictionaries>, Accessed on May 11 in 2021

<sup>33</sup> How record linkage works – CHeReL.

<https://www.cherel.org.au/how-record-linkage-works>, Accessed on May 11 in 2021

<sup>34</sup> Quality assurance - CHeReL

<https://www.cherel.org.au/quality-assurance>, Accessed on May 11 in 2021

<sup>35</sup> Technical details – CHeReL.

<https://www.cherel.org.au/how-record-linkage-works/technical-details>, Accessed on May 11 in 2021

<sup>36</sup> Apply for linked data - NSW – CHeReL.

<https://www.cherel.org.au/apply-for-linked-data>, Accessed on May 11 in 2021

<sup>37</sup> Pricing – CHeReL.

<https://www.cherel.org.au/pricing>, Accessed on May 11 in 2021

<sup>38</sup> Danielle Ní Chróinín, et al. Health-services utilisation amongst older persons during the last year of life: a population-based study. *BMC Geriatr.* 2018 Dec 20;18(1):317.

<sup>39</sup> ICES - Mission, Vision & Values.

<https://www.ices.on.ca/About-ICES/Mission-vision-and-values>, Accessed on May 11 in 2021

<sup>40</sup> ICES – FAQs.

<https://www.ices.on.ca/About-ICES/FAQs>, Accessed on May 11 in 2021

<sup>41</sup> ICES Data.

<https://www.ices.on.ca/Data-and-Privacy/ICES-data>, Accessed on May 11 in 2021

<sup>42</sup> ICES Data Dictionary.

<https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx>, Accessed on May 11 in 2021

<sup>43</sup> Working with ICES Data.

<https://www.ices.on.ca/Data-and-Privacy/ICES-data/Working-with-ICES-Data>, Accessed on May 11 in 2021

<sup>44</sup> Mira Goldberg, et al. Patterns of adjuvant care and outcomes of elderly women with stage I breast cancer after breast-conserving surgery: a population-based analysis. *Breast Cancer Res Treat.* 2019 Aug;176(3):657-667.

<sup>45</sup> Jennifer L St Sauver, et al. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epidemiol.* 2012 Dec;41(6):1614-24.

<sup>46</sup> Walter A. Rocca, et al. History of the Rochester Epidemiology Project: Half a Century of Medical Records Linkage in a US Population. *Mayo Clin Proc.* 2012 Dec; 87(12): 1202–1213.

<sup>47</sup> For Researchers | Rochester Epidemiology Project, <https://rochesterproject.org/for-researchers/>, Accessed on May 11 in 2021

<sup>48</sup> Jennifer L St Sauver, et al. Use of a Medical Records Linkage System to Enumerate a Dynamic Population Over Time: The Rochester Epidemiology Project. *Am J Epidemiol.* 2011 May 1;173(9):1059-68.

<sup>49</sup> William V. Bobo, et al. Frequency and predictors of the potential overprescribing of antidepressants in elderly residents of a geographically defined U.S. population. *Pharmacol Res Perspect.* 2019 Jan 23;7(1):e00461.

表1 データベース間の比較

	データベース名	国	設立年	運営組織	リンケージ方法	利用料	特徴
1	SEER-Medicare linked database	米国	1991年	国立研究所	DL	有料	がん患者の追跡
2	Health and Retirement Study (HRS)	米国	1992年	大学	DL (一部PL)	無料	生活習慣調査
3	Clinical Practice Research Datalink (CPRD)	英国	2012年	官公庁	DL	有料	豊富な医療情報
4	Survey of Health, Ageing and Retirement in Europe (SHARE)	EU	2002年	国立研究所	DL	無料	質問表含む
5	The Centre for Health Record Linkage (CHeReL)	オーストラリア	2006年	独立公共機関	PL	有料	死亡情報含む
6	ICES	カナダ	1992年	NPO	PL	有料	州内悉皆データ
7	Rochester Epidemiology Project medical records - linkage system (REP)	米国	1966年	医療提供機関	PL	無料	剖検情報含む

DL: Deterministic Linkage, PL: Probabilistic linkage

表 2 SEER-Medicare linked database

データベース・組織名	SEER-Medicare linked database
所属国・州	アメリカ合衆国
設立年	1991 年
種別	リンケージデータベース
データベース管理機関	アメリカ国立がん研究所 (National Cancer Institute, NCI)
資金提供機関	アメリカ合衆国保健福祉省 (Department of Health and Human Services)
データ対象	SEER: 米国人人口の約 35%が居住する 21 州のがん患者のデータベース メディケア: 65 歳以上の米国民の約 94%が受給する公的医療保険制度。長期的な障害や疾患を持つ 65 歳未満も参加 65 歳以上の SEER 登録者の約 95%がメディケアの利用情報とリンケージされている
データ期間	1991 年～現在
データコンテンツ	SEER: がんの診断部位、診断日、腫瘍の特徴 (組織構造、グレード、ステージ)、初発腫瘍数、生存状態、人口動態学的情報 (診断年齢、性別、人種、民族など)、National Center for Health Statistics の保有する死因情報 メディケア: 性別、人種、居住州、パート A～D の参加情報、診断コード (ICD コード)、短期・長期の入退院情報、医療・介護サービスの請求情報、パート D の処方情報など (提供データには社会保障番号、氏名、誕生日などの個人情報含まれない)
データ収集方法	SEER (Surveillance, Epidemiology and End Results) のがん患者データを管理する NCI が、メディケアの参加者情報・請求情報をメディケア・メディケイド・サービスセンター (Centers for Medicare & Medicaid Services, CMS) から収集し、リンケージして提供する

リンケージできる外部データ	外部データと個人レベルのリンケージが可能。NCI が個人を特定する情報をリンケージ先の組織に受け渡さない形でデータを授受し、リンケージデータを生成する
リンケージ方法	<b>Deterministic Linkage</b> 氏名、性別、生年月日、死亡日、社会保障番号を照合する リンケージ完了後、氏名、社会保障番号はファイルから消去される
データ利用方法	研究目的を記載した申請書の提出と、NCI と SEER による承認が必要。要求するデータセットの内容に応じて料金を請求する。 ※アメリカ国外の研究者には公開していない
主な研究例	SEER のがん診断情報とメディケアの人口統計データをリンケージし、がんの初診断時のステージの傾向と、人種、予防サービスの利用頻度の関連性を検証

表 3 Health and Retirement Study (HRS)

データベース・組織名	Health and Retirement Study (HRS)
所属国・州	アメリカ合衆国
設立年	1992 年
種別	研究プロジェクト
データベース管理機関	ミシガン大学社会調査研究所 (Institute for Social Research (ISR) at the University of Michigan)
資金提供機関	アメリカ国立老化研究所 (National Institute on Aging, NIA) 社会行動研究部門 (Division of Behavioral and Social Research, BSR) アメリカ合衆国社会保障局 (Social Security Administration)
データ対象	51 歳以上の米国在住者累計 37,000 人以上 (現在のパネルは約 20,000 人)
データ期間	1992 年～現在 (2 年毎にフォローアップ)
データコンテンツ	コア・インタビュー (人口動態学的情報、健康状態、家族構成、住居、経済状況など) 体力・バイオマーカー測定、心理社会的調査
データ収集方法	参加者を直接調査して収集する。参加者の半数には対面のインタビュー、体力・バイオマーカー測定を行い、残りの半数は電話によるコア・インタビューのみを実施する 参加者の生存情報は HRS の調査と National Death Index (NDI) との突合で確認する。死亡確認時は配偶者や子供にインタビューを行い、死亡した参加者の医療費、家族との交流、資産の処分などを調査する
リンケージできる外部データ	社会保障、メディケア、復員軍人援護局 (Veteran' s Administration)、雇用年金、National Death Index の死亡情報 (2014 年まで)、アメリカ合衆国退役軍人省 (Veterans Affairs) の医療記録などの行政データ

リンケージ方法	<p>リンケージする行政データに応じて異なる</p> <ul style="list-style-type: none"> <li>・メディケア・メディケイド・サービスセンター（Centers for Medicare &amp; Medicaid Services, CMS）のデータとの突合 では、HRS が第三者機関である Acumen にデータを送信し、Acumen が HRS と CMS の記録を突合してファイルを生成する</li> <li>・退役軍人省のデータとの突合 では、HRS のファイルが VA Information Resource Center (VIReC) に送信され、個人情報の Probabilistic linkage が行われる</li> </ul>
データ利用方法	<p>HRS の基本的な調査結果は HRS サイトのユーザーアカウントを作成すれば無料で利用可能。アカウントの作成には、氏名、国籍、住所、所属組織名などの入力が必要。</p> <p>認知力測定・バイオマーカーの結果や、行政データとの個別リンケージなど、機密性が高いデータへのアクセスには追加的な申請が必要になる。</p>
主な研究例	<ul style="list-style-type: none"> <li>・HRS 参加者のアルコールの消費量情報と NDI の死亡データをリンケージし、アルコール摂取と高齢者の死亡率の相関性を調査</li> <li>・HRS 参加者の情報、NDI の死亡データ、メディケアの請求データをリンケージし、参加者の死亡前 6 ヶ月間の医療支出における人種的、民族的傾向の差異を調査</li> </ul>

表4 Clinical Practice Research Datalink (CPRD)

データベース・組織名	Clinical Practice Research Datalink (CPRD)
所属国・州	英国 (United Kingdom)
設立年	2012年 (前身である Value Added Medical Products (VAMP)の設立は1987年)
種別	リンケージデータベース、リサーチサービス
データベース管理機関	イギリス医薬品・医療製品規制庁 (Medicines and Healthcare products Regulatory Agency, MHRA)
資金提供機関	イギリス医薬品・医療製品規制庁 (Medicines and Healthcare products Regulatory Agency)、 イギリス国立健康研究所 (National Institute for Health Research) (両機関ともイギリス保健省 (Department of Health and Social Care) の下部機関) Wellcome Trust, Medical Research Council, NIHR Health Technology Assessment programme, Innovative Medicines Initiative, Technology Strategy Board, Seventh Framework Programme EU, and various universities, contract research organizations and pharmaceutical companies
データ対象	英国に在住記録がある患者 5000万人 (現在の居住者では1600万人)
データ期間	1987年～現在
データコンテンツ	(一次診療データ) 診断・症状、医薬品支出、ワクチン接種歴、臨床検査、専門家による介護記録 患者の人口統計的特性 (定期的にリンケージを行う外部の医療・行政データベース) 入退院、救急医療、治療アウトカム、死亡記録、がん診断、メンタルヘルスなどの医療情報 患者の社会経済学的情報
データ収集方法	全国の一般診療医 (1900以上) が提供
リンケージできる外部デー	英国の医療・行政データベースとのリンケージが可能

タ	
リンケージ方法	<p><b>Deterministic Linkage</b></p> <p>保健省の下部機関である NHS Digital が行う。データベース内の国民保険番号 (NHS Number)、郵便番号、生年月日、性別などの個人情報を用いて、8 段階の <b>Deterministic Linkage</b> によって個人を突合する。CPRD が保有し研究者に渡すデータは匿名化されており、個人の識別情報は消去されている</p>
データ利用方法	<p><b>Independent Scientific Advisory Committee (ISAC)</b>によるプロトコル申請書の承認、ライセンス料金の支払いが必要</p>
主な研究例	<p>糖尿病患者の低血糖症による入院傾向について、高齢患者とそれ以外の年齢層の患者群を調査</p> <p>CPRD の診断情報、人口動態情報、処方情報と、入院患者のデータベース <b>Hospital Episode Statistics</b> をリンケージして、低血糖症の入院患者を分析</p>

表 5 Survey of Health, Ageing and Retirement in Europe (SHARE)

データベース・組織名	Survey of Health, Ageing and Retirement in Europe (SHARE)
所属国・州	ヨーロッパ連合
設立年	2002 年
種別	多国間パネルデータベース、管理機関
データベース管理機関	Munich Center for the Economics of Aging (MEA) (マックス・プランク社会法・社会政策研究所 (Max Planck Institute for Social Law and Social Policy) の 1 部門)
資金提供機関	欧州委員会 (European Commission) アメリカ国立老化研究所 (US National Institute on Aging) ドイツ連邦教育研究省 (German Federal Ministry for Education and Research) マックス・プランク研究所 (Max Planck Society for the Advancement of Science)
データ対象	ヨーロッパの 27 カ国とイスラエルに在住する 50 歳以上の対象者 14 万人
データ期間	2004 年～現在 (2 年毎にフォローアップ)
データコンテンツ	質問票 (年齢、性別、既往歴、入院歴、ワクチン接種歴、家庭、雇用、経済状態など) 体力・バイオマーカー測定 (身長、体重、握力、心肺機能、歩行能力など) 血液検査 (12 カ国のみ実施)
データ収集方法	参加者を直接調査する。質問票、体力・バイオマーカー測定

リンケージできる外部データ	<p>3 か国の行政データとのリンケージが確立済</p> <ul style="list-style-type: none"> <li>・ SHARE-RV: ドイツ年金基金 (German Pension Fund, DRV) の行政データ</li> <li>・ REGLINK-SHAREDK: デンマーク統計局 (Statistics Denmark)、Danish Health Data Authority の行政データ</li> <li>・ Linkage SHARE NL: オランダ統計局 (Statistics Netherlands, CBS) の行政データ</li> </ul>
リンケージ方法	<p>Deterministic Linkage (個人番号を利用)</p> <ul style="list-style-type: none"> <li>・ SHARE-RV: 社会保障番号 (Social Security Number) を使用</li> <li>・ REGLINK-SHAREDK: 個人番号 (Central personal identification number) を使用</li> </ul>
データ利用方法	<p>利用表明書の提出と承認が必要。無料で利用可能</p> <p>研究利用以外の目的ではアクセスできない</p>
主な研究例	<p>65 歳以上の多剤併用における関連因子について、18 カ国を横断調査</p> <p>質問票で取得した 5 種類以上の多剤併用の有無と、年齢・性別・運動習慣などの変数を個人レベルで回帰分析した</p>

表 6 The Centre for Health Record Linkage (CHeReL)

データベース・組織名	The Centre for Health Record Linkage (CHeReL)
所属国・州	ニューサウスウェールズ州、オーストラリア首都特別地域（オーストラリア）
設立年	2006 年
種別	データリンケージユニット
データベース管理機関	The Centre for Health Record Linkage (CHeReL)
資金提供機関	ニューサウスウェールズ州保健省（The NSW Ministry of Health）
データ対象	ニューサウスウェールズ州およびオーストラリア首都特別地域に行政・医療記録がある 900 万人以上
データ期間	1963 年～現在
データコンテンツ	主に CHeReL 内部のリンケージシステム Master Linkage Key に含まれる行政・医療データセットを提供する 入退院記録、緊急診療記録、感染症記録、がん患者登録簿、出生・死亡記録など
データ収集方法	管轄地域の医療・研究機関、行政組織から取得する (2016 年以前) CHeReL はデータを保管せず、研究リクエストに応じてデータを保有する参加機関に問い合わせ、研究ごとにリンケージを行っていた。この方法は機密性に優れる一方、データ抽出の煩雑化が指摘されていた (2017 年以降) CHeReL は参加機関から定期的にデータを収集し、個人を特定する情報を分離して保管する。リクエストに応じて研究者にデータ提供を行う
リンケージできる外部データ	Master Linkage Key のデータセットに含まれない外部の研究機関のデータ、行政データとのリンケージが可能

リンケージ方法	<p>主に Probabilistic linkage を用いる</p> <p>個人の突合には Choicemaker (ソフトウェア) による個人情報の Probabilistic linkage を行う (偽陽性率 0.5%を目標として設計)</p> <p>研究者に提供されるデータには CHeReL が研究毎に生成する PPN (Project specific Person Number for each individual) を付与し、個人を特定する氏名などの情報は含まれない</p>
データ利用方法	<p>研究プロトコルと申請書の承認、ライセンス料金の支払いが必要</p>
主な研究例	<p>2007年にニューサウスウェールズ州で死亡した 70 歳以上の 34,556 名の入院記録、がん登録情報、死亡記録をリンケージ。最終生存年における医療サービス利用の特徴を調査</p>

表7 ICES

データベース・組織名	ICES
所属国・州	オンタリオ州（カナダ）
設立年	1992年
種別	非営利研究機関
データベース管理機関	ICES
資金提供機関	オンタリオ州保健・介護省（Ministry of Health and Long-Term Care） 民間の資金提供は受けていない
データ対象	オンタリオ州の在住者約1300万人
データ期間	1986年～現在
データコンテンツ	医療保険請求、医薬品請求、入退院、在宅ケア、介護サービスの利用情報 がん、脳卒中、心臓病などの特定疾患患者の記録 慢性疾患患者のコホート 手術記録 Canadian Community Health Survey などの大規模な健康調査結果 ICES primary data collection projects の詳細な臨床データ 国勢調査、人口動態学的データ
データ収集方法	オンタリオ州の公的資金を受けた医療・公的機関から、医療サービスデータ、人口動態学的データを収集
リンケージできる外部データ	

リンケージ方法	<p><b>Probabilistic Linkage</b></p> <p>個人の突合には <b>AutoMatch</b> ソフトウェアを使用した個人情報の <b>Probabilistic linkage</b> を行う。</p> <p><b>ICES</b> はデータを取得後、氏名などの個人情報を消去し、<b>Ontario health card number</b> を暗号化して固有の識別番号 (<b>ICES Number, IKN</b>) を割り当てる。<b>ICES</b> のデータを利用する研究者は <b>IKN</b> に暗号化されたデータにしかアクセスできない。</p>
データ利用方法	研究プロトコルの承認とライセンス料金の支払いが必要
主な研究例	2010 年から 2016 年の間にステージ I 乳がんの診断を受けた 65 歳以上の患者を対象に、内分泌療法、放射線療法、補助的治療のアウトカムの差異を調査

表 8 Rochester Epidemiology Project medical records - linkage system (REP)

データベース・組織名	Rochester Epidemiology Project medical records - linkage system (REP)
所属国・州	ミネソタ州オルムステッド郡 (米国)
設立年	1966 年
種別	医療データベースリンケージシステム
データベース管理機関	Olmsted Medical Center, Mayo Clinic, and Rochester Family Medicine Clinic
資金提供機関	アメリカ国立衛生研究所 (National Institutes of Health, NIH) アメリカ国立老化研究所 (National Institute on Aging)
データ対象	オルムステッド郡の在住者 502,820 名 (2012 年時点)
データ期間	1966 年～現在
データコンテンツ	患者の人口統計学的情報 (氏名、性別、生年月日、住所) Mayo Clinic patient number などの医療提供者固有の患者個人番号 疾病コード (ICD-9)、診療行為コード (CPT) 医薬品処方情報 死亡情報 (死亡日、死因) オルムステッド郡の剖検情報
データ収集方法	参加医療機関が REP に提供 ・ミネソタ州、ウィスコンシン州の 50 以上の医療提供機関 (2012 年時点) ・ミネソタ州 (死亡証明書) ・アメリカ疾病予防管理センター (National Death Index を取得)
リンケージできる外部データ	外部の医療データが REP に参加する場合は、REP の研究チームがリンケージを行う

リンケージ方法	<p>Probabilistic linkage</p> <p>(2002 年以降) SAS software package (SAS Institute, Inc., Cary, North Carolina) の Soundex phonetic filing system を採用。氏名、性別、生年月日、社会保険番号を合計 20 段階の方法で突合。その後、不明確な組み合わせは REP のテクニカルチームが手動で確認する。</p>
データ利用方法	<p>Mayo Clinic を通して"REP Collaborator"役職の指名と、Mayo Clinic、Olmsted Medical Center (OMC)、Institutional Review Boards (IRBs)による研究プロトコルの承認が必要</p>
主な研究例	<p>65 歳以上の高齢者を対象に、抗うつ剤の過剰処方に関する予測因子を調査</p> <p>処方情報に対して、併発疾患、居住環境、医療サービス利用を説明変数とする回帰分析を行った</p>