2026 年 3 月 31 日

# ＜研究課題＞プロブレムリストや高齢者総合機能評価等のテキストデータをデータベース化するプログラム開発とそれを用いた臨床転帰および医療費との関連に関する分析研究

代表研究者　国立長寿医療研究センター医療経済研究部副部長　大西　丈二

（老年内科医長、地域医療連携室長、医療情報室長）

**【抄録】**

電子カルテには高齢者総合機能評価（CGA）等の重要情報がテキスト・テンプレート形式で保存され二次利用が困難である。国立長寿医療研究センター病院の電子カルテ帳票（入退院歴・退院サマリー・CGA・退院支援スクリーニング・リハビリテーション処方）を入院エピソード単位のデータベースへ構造化するプログラムを Python と DuckDB で開発した。2023～2025 年度の 15,950 入院（9,366 名）を分析し、CGA 捕捉率 94.3%、担当医特定率 99.8%を得た。機能障害を有する在院 7 日以上の入院 6,214 件中、リハ処方のない「ギャップ」は 544 件（8.8%）で、同群の 30 日再入院率は 15.4%とリハ処方群の 10.9%より高かった。入院からリハ処方までは中央値 0 日、2 日以内の早期処方率は 82.0%から 87.3%へ改善した。プログラムは診療の質指標の継続的算出を可能する有用性が示唆された。

## １．研究の目的

### 1-1 背景

高齢入院患者の診療では、プロブレムリスト、総合機能評価（Comprehensive Geriatric Assessment; CGA）、退院支援スクリーニング等の評価が日常的に行われているが、多くは非構造化テキストや電子カルテのテンプレートとして保存され、施設横断的な分析や質改善への二次利用が十分に進んでいない。CGA は入院高齢者の在宅生存を高めるエビデンスを有する[1, 2]一方、老年症候群・機能障害の記載は診断コードに反映されにくく、非構造化データの活用が課題とされている[3-6]。英国の Electronic Frailty Index（eFI）[7]や Hospital Frailty Risk Score[8]のように標準指標を電子カルテへ実装する方式に対し、わが国では生活機能を含む標準規格が整備途上にある。

### 1-2 目的

現行の評価・記録方法を変更せずに、既存の診療記録から構造化データベースを自動生成するプログラムを開発し、それを用いて臨床転帰および医療費との関連を分析するための基盤を構築することを目的とした。

## ２．研究方法と経過

### 2-1 デザイン・対象・倫理

本研究は、診療記録等を用いて、介入のない後ろ向き観察研究として実施した。国立長寿医療研究センター病院の全診療科入院エピソードを対象とし、主解析は帳票の整備状況が揃う 2023～2025 年度退院分とした。

### 2-2 データ抽出

電子カルテ HOPE EGMAIN-GX およびデータウェアハウス HOPE DWH-GX（富士通ジャパン株式会社）から入退院歴、患者基本情報、退院サマリー、リハビリテーション（以下、リハ）処方、および当院独自テンプレートとして CGA、退院支援スクリーニングの各帳票を年度別 CSV として出力し、外部ネットワークから切り離したスタンドアロン解析機（Ubuntu 24.04）に格納した。

### 2-3 データベース化プログラムの開発

Python（pandas）と DuckDB を用い、多段階の ETL（抽出・変換・格納）として実装した。

第 1 段階（データマート化）では、帳票ごとに文字コードとファイル形式を自動判別して読み込み、列名・日付・全角半角の正規化、HOPE GX ローカル病名コード等の表記ゆれ

処理を行い、列指向形式（Parquet）へ変換した。

第 2 段階（データキューブ化）では、患者 ID と入院日から入院エピソード ID を生成し、同日入退院や重複行を規則的に統合したうえで、全帳票をエピソード単位に連結し DuckDB データベースに格納した。担当医は退院サマリー作成医を優先し、欠損時は入退院歴の主治医で補完した。リハ処方は処方日を用いて入院期間内で紐づけ、入院前 7 日以内の事前オーダも含めた。回復期リハビリテーション病棟の入院例においては、入院 28 日前以内のオーダも有効とした。

第 3 段階（集計・分析）では、科別・年度月別のクロス集計、リハ処方関連指標、転帰指標を自動出力するスクリプトを作成した。コード作成には生成 AI（Anthropic 社 Claude）を活用し [9,10]、実データによる検証と修正を反復して品質を確保した。

### 2-4 分析

CGA（12 項目）および退院支援スクリーニングの該当状況を機能障害情報として用い、（a）機能障害項目別のリハ処方率、（b）科別・年度別リハ処方率、（c）リハ適応となりうる機能障害（ADL 低下・転倒リスク・嚥下障害等）を有しながらリハ処方のない「ギャップ」該当例の頻度と特徴（短期入院の影響を除くため在院 7 日以上に限定）、（d）機能障害・リハ処方の有無による 4 群の転帰（死亡退院、全期間の入退院歴から算出した 30 日・90 日再入院）、（e）入院からリハ処方までの待機日数（事前オーダは 0 日、早期処方率＝2 日以内）を集計した。医療行為と転帰の関連分析はリハ処方を代表例として検討した。群間比較には $\chi^2$ 検定を用いた。

### 2-5 倫理的配慮

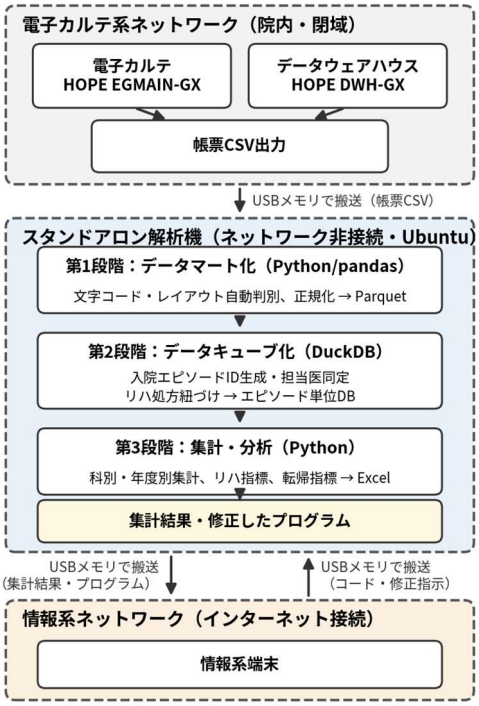
本研究は、当センター倫理・利益相反委員会の承認（No.1777）のもと、個人情報保護関連法令・ガイドラインおよび施設内規程に従い実施した。

## ３．研究の成果

### 3-1 解析プログラムの開発

解析環境の整備、帳票仕様の把握と試験抽出、プログラムの開発・実データを用いた検証と修正、集計・分析の順に進め、図 1 の解析プログラムを開発した。

**図 1. 解析の手順**




### 3-2 データベース化の性能

2023～2025 年度の全科入院のべ 15,950 例（9,366 名）が対象とされた。対象特性を表 1 に示す。CGA 捕捉率 94.3%、退院支援スクリーニング捕捉率 98.9%、担当医特定率 99.8%であり、手作業のカルテレビューに比して大幅な効率化を達成した。

**表 1　対象特性（2023～2025 年度）**

| 項目 | 値 |
|---|---|
| 入院エピソード数 | 15,950 |
| 2023／2024／2025 年度 | 5,004／5,302／5,644 |
| 患者数 | 9,366 |
| 年齢，中央値［IQR］ | 80［74–85］歳 |
| 女性 | 51.1% |
| 在院日数，中央値［IQR］ | 10［2–26］日 |
| CGA 記載あり | 94.3% |
| 退院支援スクリーニングあり | 98.9% |
| リハ処方あり（補正後） | 54.9%（55.3%） |
| 死亡退院 | 2.4% |

IQR：四分位範囲。補正後：回復期リハ病棟自科入棟の補正を含む。

### 3-3 機能障害

機能障害として、CGA 記載例のうち ADL 低下は 37.0%、もの忘れ 29.3%、うつ症状 3.2%、意欲低下 11.2%、転倒リスク 42.1%、嚥下障害 12.7%、排泄障害 25.0%、難聴 16.6%、5 種類以上の服薬 54.8%で認められた。

### 3-4 退院支援スクリーニング

退院支援が必要とされる例として、スクリーニング記載例のうち、悪性腫瘍・認知症又

は誤嚥性肺炎等の急性呼吸器感染症のいずれか 18.4%、緊急入院 41.6%、介護保険の被保険者であり未申請（又は、要介護度の見直しが必要な状態）13.0%、コミュニケーションに特別な技術が必要な障害を有する 1.5%、強度行動障害の状態 0.1%、医療保険未加入者又は生活困窮者 0.5%、入院前に比べ ADL が低下し退院後の生活様式の再編が必要 40.2%、排泄に介護を要する 30.2%、必要な養育又は介護を十分に提供できる状況にない 10.3%、退院後に医療処置が必要 7.2%、入院治療を行っても長期的な低栄養状態になることが見込まれる 4.4%、家族に対する介助や介護等を日常的に行っている児童等 0.0%、患者の意思決定支援等を行うに当たり家族や親族との連絡が困難 0.0%、児童等の家族から介助や介護等を日常的に受けている 0.1%が該当した。

### 3-5 リハ処方

リハ処方（補正後）は、65-74 歳では 41.2%、75-84 歳では 54.5%、85-94 歳では 73.8%、95 歳以上では 82.3%、全体では 55.3%で行われていた。リハ科以外では、老年内科 93.4%、整形外科 86.8%、糖尿病・内分泌内科 86.2%の順で多く処方されていた。

入院からリハ処方までの待機日数は中央値 0 日で、平均 1.6 日から 1.3 日へ、早期処方率 82.0%から 87.3%へ、90 パーセンタイル値 5 日から 3 日へと年度を追って改善していた（表 2）。

**表 2　入院からリハ処方までの待機日数の年度推移**

| 年度 | n | 平均±SD | 中央値 | p90 | 早期率 |
|---|---|---|---|---|---|
| 2023 | 2,709 | 1.6±3.5 | 0 日 | 5 日 | 82.0% |
| 2024 | 2,919 | 1.4±3.8 | 0 日 | 4 日 | 85.5% |
| 2025 | 3,127 | 1.3±4.7 | 0 日 | 3 日 | 87.3% |

早期率：入院 2 日以内の処方（入院前 7 日以内の事前オーダは 0 日として算入）。

### 3-6 プロブレムリスト

プロブレムリストを退院サマリーから集め、ICD-10 および独自の病名リストに振り直す対比表および変換プログラムを作成した。しかし退院サマリーでは退院時病名欄に記載される病名以外で、体系的なプロブレムリストが記載されている例が極めて少なく、全入院患者を対象とした分析を行うには至らなかった。

### 3-7 機能障害および退院支援スクリーニングとリハ処方

機能障害項目別のリハ処方率は、意欲低下 86.4%、ADL 低下 85.9%、嚥下障害 85.6%と高かった一方、5 種類以上の服薬は 65.9%と低かった。在院 7 日以上かつリハ適応となりうる機能障害を有する 6,214 件のうち、リハ処方のない例は 544 件（8.8%）あり、科別では消化器内科 13.4%、脳神経内科 8.3%でリハ処方が行われていない率が高く、診療科間差が有意に認められた（$p < 0.001$）。

退院支援スクリーニング該当とリハ処方率については、長期的な低栄養状態が見込まれる例で 86.0%、次いでコミュニケーションに特別な技術が必要な障害を有する例で 85.7%と多く、退院後に医療処置が必要な例で 62.7%、次いで介護保険未申請（又は要介護度の見直しが必要）な例で 62.8%と少なかった。

### 3-8 リハ処方と転帰および再入院

障害を持つ例について、リハ処方有群と無群の転帰を表 3 に示す。障害有・リハ無群の軽快・治癒・寛解での退院は 79.2%、死亡退院率は 4.8%で、障害有・リハ有群ではそれぞれ 77.2%、4.6%であった。

障害有・リハ無群の 30 日・90 日再入院率は、障害有・リハ有群と比べ有意に高かった（それぞれ $p = 0.002$、$p = 0.009$）。

**表 3　機能障害を有する例のリハ処方別の転帰（在院 7 日以上）**

| 群 | n | 在院日数中央値 | 30 日再入院 | 90 日再入院 |
|---|---|---|---|---|
| 障害あり・リハあり | 5,670 | 28 日 | 10.9% | 20.5% |
| 障害あり・リハなし | 544 | 13 日 | 15.4% | 25.4% |

再入院は全期間の入退院歴から算出（死亡退院は分母から除外）。死亡退院率はリハあり 4.6%、リハなし 4.8%。

## 4．今後の課題

退院支援スクリーニング項目は診療報酬上の加算で規定されているが、CGA は総合評価加算の規定を満たし、当院のユニークな評価である [11,12]。現在、厚生労働省によって 3 文書 6 情報を共有する、電子カルテ情報共有サービスが進められているが、機能障害は診療上、主要な情報項目であるものの、必須の共有項目とされていない。本研究はその中で、CGA および退院支援スクリーニング項目を他項目と合わせ分析できる基盤を構築したことがユニークな点である。本研究ではそれらとリハ処方とを分析することによって、機能低下例のうちリハ処方実績を分析することにより、必要な例に必要なリハが行われているか、

適切な時期にリハが行われているかを検討することができた。急性期入院に伴う機能低下[13]は再入院と関連し[14]、早期リハの意義が報告されていることから[15]、これらは診療の質指標の候補となる[16]。

今回の研究でリハ処方と分析を行ったのと同様に、診断・治療内容や、多職種チームの関与、医療費等の分析を進めることができる基盤を構築することが、本研究の主要な成果である。

本研究の限界を以下に述べる。リハ処方との関連分析については、まずすべての分析は後ろ向き研究であり、転帰との関連等、因果関係の断定はできないことである。第二は、リハ処方は実施そのものを保証せず、実施記録との突合が必要である。第三に、申請時に計画したプロブレムリストを用いた分析は、上記 3-6 で記載した通り、十分に行うことができず、今後の課題として残った。第四に、同じく申請時に計画した医療費・原価分析は、DPC データ[17]・レセプトとの連結により継続実施中であるが、解析を研究期間内に終えることができず、研究を継続する。第五に、単一施設の成果であり、他施設の電子カルテ帳票への適用と多施設展開により汎用性を検証する必要がある等がある。共通データモデル[18]への準拠も今後の検討課題である。

## ５．研究成果の公表方法

開発したプログラムの一部は国立長寿医療研究センター医療情報室サイト（ https://www.ncgg.go.jp/hospital/overview/organization/his）にて公開する。高齢者機能評価の標準化とデータ二次利用推進の基礎資料とする。ほか、本研究の成果は、国内・国外の学会発表を行い、実データを用いた分析を進め、国外の論文誌として投稿する予定である。

## 参考文献


1) Stuck AE et al. Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet 1993;342(8878):1032-1036.
2) Ellis G, et al. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database Syst Rev 2017;9(9):CD006211.
3) Kharrazi H, et al. The value of unstructured electronic health record data in geriatric syndrome case identification. J Am Geriatr Soc 2018;66(8):1499-1507.
4) Anzaldi LJ, et al. Comparing clinician descriptions of frailty and geriatric syndromes using electronic health records: a retrospective cohort study. BMC Geriatr 2017;17(1):248.
5) Chen T, et al. Identifying vulnerable older adult populations by contextualizing geriatric syndrome information in clinical notes of electronic health records. J Am Med Inform Assoc 2019;26(8-9):787-795.
6) Osman M, et al. The use of natural language processing for the identification of ageing syndromes including sarcopenia, frailty and falls in electronic healthcare records: a systematic review. Age Ageing 2024;53(7):afae135.
7) Clegg A, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing 2016;45(3):353-360.
8) Gilbert T, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet 2018;391(10132):1775-1782.
9) Kawazoe Y, et al. A clinical specific BERT developed using a huge Japanese clinical text corpus. PLoS One 2021;16(11):e0259763.
10) 宇野裕, 石井亮, 柴田大作, ほか. 大規模言語モデルを利用した治療経過サマリの作成支援アプリの試作. 2024 年度人工知能学会全国大会（第 38 回）論文集; 2024.
11) 鳥羽研二 監修, 長寿科学総合研究 CGA ガイドライン研究班 著. 高齢者総合的機能評価ガイドライン. 東京: 厚生科学研究所; 2003.
12) 長寿医療研究開発費「高齢者総合機能評価（CGA）ガイドラインの作成研究」研究班, 日本老年医学会, 国立長寿医療研究センター編. 高齢者総合機能評価（CGA）に基づく診療・ケアガイドライン 2024. 東京: 南山堂; 2024.
13) Covinsky KE, et al. Hospitalization-associated disability: "She was probably able to ambulate, but I'm not sure". JAMA 2011;306(16):1782-1793.
14) Greysen SR, et al.. Functional impairment and hospital readmission in Medicare seniors. JAMA Intern Med 2015;175(4):559-565.
15) Momosaki R, et al. Effect of early rehabilitation by physical therapists on in-hospital mortality after aspiration pneumonia in the elderly. Arch Phys Med Rehabil 2015;96(2):205-209.
16) Mitsutake S, et al. Predictive validity of hospital-associated complications of older people identified using Diagnosis Procedure Combination data from an acute care hospital in Japan: observational study. JMIR Aging 2025;8:e68267.
17) Yasunaga H. Real world data in Japan: Chapter II. The Diagnosis Procedure Combination database. Ann Clin Epidemiol 2019;1(3):76-79.
18) Hripcsak G, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578.

March 31, 2026

# Development of a program to structure free-text and template-based clinical records, including problem lists and comprehensive geriatric assessment, into a database, and analysis of associations with clinical outcomes and healthcare costs

Primary Researcher: Joji Onishi, MD, PhD, MPH — Deputy Director, Department of Healthcare Economics Research (concurrently Chief, Department of Geriatric Internal Medicine), National Center for Geriatrics and Gerontology, Obu, Aichi, Japan

**Abstract**

Important information for older inpatients, such as comprehensive geriatric assessment (CGA), is stored as free text or templates in electronic health records, which hinders secondary use. We developed a program in Python and DuckDB that structures electronic health record extracts from the National Center for Geriatrics and Gerontology Hospital (admission histories, discharge summaries, CGA, discharge-support screening, and rehabilitation orders) into an episode-level database. Analysis of 15,950 admissions (9,366 patients) in fiscal years 2023–2025 yielded a CGA capture rate of 94.3% and an attending-physician identification rate of 99.8%. Among 6,214 admissions of seven days or longer with functional impairment, 544 (8.8%) lacked rehabilitation orders (“gap” cases); their 30-day readmission rate was 15.4%, higher than the 10.9% in the group with rehabilitation orders. Median time from admission to rehabilitation order was 0 days, and the early-ordering rate (within two days) improved from 82.0% to 87.3%. The program demonstrated its usefulness in enabling the continuous calculation of quality indicators for medical care.