

厚生労働科学研究費補助金（長寿科学政策研究事業）
「訪問系サービスにおける LIFE の活用に向けた評価指標の開発に資する研究」
分担研究報告書（令和 7 年度）

LIFE データ・データクリーニング

研究分担者 松本 伸哉 （島根大学 医学部 環境保健医学講座）
研究分担者 西岡 祐一 （奈良県立医科大学 公衆衛生学講座）
研究分担者 山口 佳小里（国立保健医療科学院 医療・福祉サービス研究部）
研究代表者 赤羽 学 （国立保健医療科学院 医療・福祉サービス研究部）

研究要旨

2021 年度より運用が開始された科学的介護情報システム（LIFE）データは項目数が多く、提出の重複や入力エラーを含む可能性があり、医療・介護レセプトのような厳格なチェック機構も十分ではないため、そのままでは分析に適さないという課題がある。本研究では、第三者提供された LIFE データを分析可能な形に整備することを目的とした。

方法として、2021 年 4 月から 2024 年 3 月までの LIFE データを用い、データウェアハウスを構築してデータクリーニングを実施した。リアルワールドデータの整理概念である「事象」と「状態」に基づき、「誰に」「いつ（必要に応じて何を）」という観点から同一事象を特定し、重複データの整理および一意性の確保を行った。また、1 レコード内に複数情報を含むファイルについては構造を分離した。

その結果、各ファイルにおいてレコード数の変動が確認され、重複整理や構造分離の影響が明らかとなった。一方で、入力ミスによる異常値や一意性判定項目自体の誤入力などがみられ、それらへの対応は今後の課題である。

本研究により、LIFE データから一意性の高いデータを抽出する手法を提示し、利用者のデータクリーニング負担の軽減および研究利用の促進に資する基盤を整備した。今後は KDB との比較分析等を通じて、さらなる検証を行う予定である。

A. 研究目的

2021 年度より科学的介護情報システム（以下「LIFE」という。）の運用が開始され、多くの介護施設・介護事業所において導入が進んでいる。また、LIFE データの第三者提供が開始されたことにより、研究機関や民間企業等によるエビデンス創出に向けた研究の推進が期待されている。

一方で、LIFE データは項目数が多岐にわた

るとともに、いくつかの運用上の課題を有している。第一に、多くの施設・事業所において既存の介護ソフトへ LIFE 提出機能を追加実装する形で対応していることから、提出タイミングが必ずしも自動化されておらず、手動提出に伴う重複提出や提出漏れの可能性がある。第二に、入力段階での誤入力十分に検証されないまま提出されている可能性がある。医療レセプトおよび介護レセプトは請求データであるため

請求上の整合性確認が行われるが、LIFE データにおいては同様の厳格な検証機構が十分とはいえない。

このような特性を踏まえ、本研究では、第三者提供された LIFE データから分析に使用可能なデータを整備することを目的とした。

B. 研究方法

本研究では、厚生労働省より提供を受けた「介護保険総合データベース」の定型データセットのうち、2021 年 4 月から 2024 年 3 月までの LIFE データを対象とした。分析基盤としてデータウェアハウス (DWH) を構築し、データ整理および集計作業を実施した。

リアルワールドデータを整理する概念として、「事象 (event)」と「状態 (state)」の枠組みを採用した。事象とは特定時点に発生した変更不可能な出来事を指し、状態とは特定時点における状況を示す。本研究では、複数回送信された可能性のあるデータについて、「誰に」「いつ」(必要に応じて「何を」)という観点から事象として再定義し、同一事象につき 1 レコードとなるよう一意化処理を行った。

「誰」に関しては、LIFE が施設単位で報告される仕様であることから、在宅サービス等において同一利用者が複数施設からサービスを受けている場合には別人として扱う仕様を前提とし、これを許容した。「いつ」については、ファイル種別ごとに事象特定に用いる項目を定義した。

なお、一部ファイルについては例外的な取り扱いを行った。既往歴情報、服薬情報、薬剤変更情報については、「何を」に該当する情報を追加して事象を特定した。また、口腔衛生管理情報、口腔機能向上サービス管理情報、その他情報については、単一レコード内に複数情報が格納されている構造であったため、これを分離し、個別の事象として再構成した。各ファイルの分割方法および事象特定項目については表 1 に示した。

C. 研究結果

上記手法に基づき、LIFE データのデータクリーニングおよび一意化処理を実施した。表 2 に、提供を受けた各ファイルのレコード数と、データクリーニング後のレコード数を示した。

多くのテーブルでは提供レコード数と概ね同水準であったが、一部のテーブルでは大幅な変動が認められた。例えば、口腔清掃実施情報については、1 回の報告で最大 6 回分の実施状況を報告可能な構造であったため、分離・再構成の結果、レコード数は 100% を超過した。また、「その他情報」については登録件数が少なかったが、必須項目ではないことがその要因として考えられる。

D. 考察

本研究では、評価日等のデータ発生日を基準として「事象」単位でデータを抽出し、一意性の高いデータセットを構築した。これらの作業によって、重複提出の影響を低減し、分析に適した構造へと整理することができたため、LIFE データの利用者が分析前に実施しなければならないデータクリーニングの負担軽減が期待される。

NDB (レセプト情報・特定健診等情報データベース) および介護 DB (介護保険総合データベース) においては、各項目の値の実態を整理したコードブックが作成・公開され^{1,2)}、研究基盤として活用されている³⁾が、LIFE データにおいても研究利用および行政利用の円滑化に資する基盤の整備を行うことができた。

一方で、入力ミスに起因する異常値への対応や、一意性判定に用いた項目自体に誤入力が存在する場合の補正については、本研究では実施していない。今後は、データ品質評価および異常値検出手法の導入が課題である。

E. 結論

本研究により、第三者提供された LIFE データから事象単位で一意性の高いデータを抽出する手法を確立し、研究利用および行政利用の円滑化に資する基盤整備を行った。

今後は、KDB 等の他データベースとの比較分析を通じて、本手法の妥当性および有用性の検証を進める予定である。

なし

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

引用文献

- 1). レセプト情報・特定健診等情報データベース (NDB) のコードブック【各項目の要約統計量及び項目値上位 50 の公表】 公立大学法人奈良県立医科大学 公衆衛生学講座
https://hpm.naramed-u.ac.jp/res_document_ndbcodebook.html
- 2). 介護DBコードブックについて 京都大学医学部 附属病院 医療情報企画部
<https://medinfo.kuhp.kyoto-u.ac.jp/document/kaigo-db-codebook/>
- 3). 森由希子, 黒田知宏, 加藤源太, 植嶋大晃, 中谷健太, 大寺祥佑. 臨床疫学に活用可能な NDB 等データセットの作成に関する研究: 介護データベースコードブック作成及び NDB データセット仕様の検討. 厚生労働行政推進調査事業補助金 (政策科学総合研究事業 (政策科学推進研究事業)) 分担研究報告書

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

表 1：各ファイルの分割と追加キー項目

テーブルNo	テーブル名	実テーブル名	分割	追加キー項目
1	利用者情報	SERVICE_USER_INFO		created_date
2	科学的介護推進情報	form_z_2021		evaluate_date
3	科学的介護推進情報（既往歴情報）	form_z_detail_medical_his_2021		injury_code, symptoms_appearance_date#NULL可能
4	科学的介護推進情報（服薬情報）	form_z_detail_medicine_2021		dispensing_date, chemical_code_type, chemical_code
5	栄養・摂食嚥下情報	form_a_2021		implementation_date
6	栄養ケア計画等情報	form_t_2021		created_date
7	口腔衛生管理情報	form_d_2021	基本情報	last_confirmed_date
			評価情報	filing_date_01
			管理情報	filing_date_02
			口腔清掃実施	oral_care_filing_date_01～06
8	口腔機能向上サービス管理情報	form_c_2021	基本情報	last_confirmed_date
			評価情報	screening_and_monitoring_date_01
			計画情報	plan_create_date
			口腔清掃実施	plan_date
9	興味関心チェック情報	form_f_2021		implementation_date
10	生活機能チェック情報	form_g_2021		evaluate_date
11	個別機能訓練計画情報	form_m_2021		created_date
12	リハビリテーション計画書（医療介護共通部分）	form2_1_2021		created_date
13	リハビリテーション計画書（介護）	form2_2_2021		created_date
14	リハビリテーション会議録（様式3情報）	form3_2021		event_date, event_start_hour#NULL可能, event_start_minute#NULL可能
15	リハビリテーションマネジメントにおけるプロセス管理票（様式4情報）	form4_2021		created_date
16	生活行為向上リハビリテーション実施計画書（様式5情報）	form5_2021		created_date
17	褥瘡マネジメント情報	form_j_2021		evaluate_date
18	排せつ支援情報	form_k_2021		evaluate_date
19	自立支援促進情報	form_l_2021		evaluate_date, plan_create_date
20	薬剤変更情報	form_n_detail_medicine_change_2021		start_date, drug_code_category, drug_code
21	薬剤変更情報（既往歴情報）	form_n_detail_medical_his_2021		injury_code, symptoms_appearance_date#NULL可能
22	ADL維持等情報	form_o_2021		adl_evaluate
23	その他情報	form_other_2021	日常動作	evaluate_date_01
			身体関連	evaluate_date_02
			認知症関連	evaluate_date_03
			自立度	evaluate_date_04

表 2：各ファイルのクリーニング前後のレコード数

テーブルNo	テーブル名	提供データレコード数	分割	クリーニング後行数	元データに対する比率
1	利用者情報	5,397,160		5,397,060	100.0%
2	科学的介護推進情報	26,806,430		23,706,850	88.4%
3	科学的介護推進情報（既往歴情報）	38,489,900		7,295,880	19.0%
4	科学的介護推進情報（服薬情報）	42,870,190		22,774,890	53.1%
5	栄養・摂食嚥下情報	11,890,740		10,725,470	90.2%
6	栄養ケア計画等情報	872,750		772,470	88.5%
7	口腔衛生管理情報	2,530,050	基本情報	2,396,330	94.7%
			評価情報	2,097,550	82.9%
			管理情報	2,045,110	80.8%
			口腔清掃実施	5,385,380	212.9%
8	口腔機能向上サービス管理情報	2,198,970	基本情報	1,997,960	90.9%
			評価情報	1,647,400	74.9%
			計画情報	1,653,760	75.2%
			口腔清掃実施	5,430,910	247.0%
9	興味関心チェック情報	2,307,560		1,931,970	83.7%
10	生活機能チェック情報	9,748,010		8,180,230	83.9%
11	個別機能訓練計画情報	10,027,270		8,441,800	84.2%
12	リハビリテーション計画書（医療介護共通部分）	5,411,630		5,202,260	96.1%
13	リハビリテーション計画書（介護）	5,410,710		5,195,610	96.0%
14	リハビリテーション会議録（様式3情報）	419,650		355,190	84.6%
15	リハビリテーションマネジメントにおけるプロセス管理票（様式4情報）	209,990		174,570	83.1%
16	生活行為向上リハビリテーション実施計画書（様式5情報）	22,230		14,670	66.0%
17	褥瘡マネジメント情報	5,380,840		4,995,390	92.8%
18	排せつ支援情報	3,617,280		3,278,940	90.6%
19	自立支援促進情報	2,386,840		2,015,340	84.4%
20	薬剤変更情報	1,385,190		694,750	50.2%
21	薬剤変更情報（既往歴情報）	811,610		176,440	21.7%
22	ADL維持等情報	7,526,750		5,937,140	78.9%
23	その他情報	3,258,900	日常動作	209,420	6.4%
			身体関連	48,170	1.5%
			認知症関連	96,440	3.0%
			自立度	112,110	3.4%