

(別紙1)

総括研究報告書

課題番号：30-26

課題名：成育データコモンズの構築

主任研究者名（所属施設） 盛一 享徳（国立成育医療研究センター研究所）

（所属・職名）小児慢性特定疾病情報室・室長

成育保健・医療に関するデータソースを一元的に検索・収集できる「成育データコモンズ」を構築することにより、既存の研究成果の二次利用を促進・効率化し、さらなる研究推進や政策立案につながると期待される。本調査にて既存の調査研究の種類には、政府統計や公的資金による調査報告書、民間シンクタンクの研究報告書などがあり、調査項目も身体的・精神的・社会的健康課題など多様であることが分かった。一方で既存のデータベースを用いて、関連の調査研究が特定されても、具体的な調査項目を知ることが困難なことも多く、新たなデータベース構築の際の課題であった。

具体的な既存データベースの二次利用の利活用として、診療報酬明細（レセプト）データを他のデータベースとの突合の際に障壁となるレセプト傷病名の取扱について検討を行った。膨大なデータを処理するには困難を極めるが、自然言語解析の技術を応用することにより、レセプト傷病名の類型化ができる可能性を見出した。

1. 研究目的

成育保健・医療分野においても他の医療分野と同様に、毎年膨大な研究・調査が提案され、多くの公的資金が投入されて施行されてきた。しかしながら、公的資金を投入されて関係者の多大な努力により収集されたデータが、二次的に最大限に利用されてきたとは言えない状況がある。また人口動態統計やレセプトデータをはじめとする政府の公的統計も、それぞれの目的にかなった

利用はされてきたが、存在しているデータを目的外の利用が最大限に行われているとは言えない。一方、成育保健・医療分野においても、新たな医療技術の開発、適正な政策・医療上の意思決定、基礎・応用研究の推進のためには、客観的なデータによる分析が必須となってきた。

それぞれの研究や調査には、それぞれの目的があり、それに合わせて個人情報保護などの規定が設けられており、一律にデー

タをプールして利用することは、困難であるが、調査研究の対象、データに含まれる変数、主体となった研究者を含めたデータセットの後方視的な登録データベース（目録）を構築することで、データの二次利用が促進されることは容易に推測される。

平成29年度子ども子育て支援推進研究調査事業における先行研究において、子どもの社会的・医学的・精神的環境を明らかにする目的をもって、目録作りを行ったが、この成果を生かし、本研究は継続的に研究基盤を整備することを主目的とする。

既存のデータベースは単体での二次利用には限界があるため、相互のリンケージが期待されている。そこで成育保健・医療分野における巨大データベースの一つである小児慢性特定疾病（小慢）登録データベースと診療情報明細（レセプト）データベースとのリンケージの可能性についても合わせて検証を行う。とくにレセプトデータは取扱が難しいことから、小慢登録データベースとのリンケージを視野に置いたレセプトデータの検証を行う。

2. 研究組織

盛一 享徳（国立成育医療研究センター）

須藤 茉衣子（国立成育医療研究センター）

3. 研究成果

本年度は、次の2つの課題について研究を行った。

1) 成育データコモンズの構築を目的としたデータベース調査

本研究に関連する調査研究について検討を行い、5,625件（2018年11月時点）が該当した。調査研究の種類は、政府統計、学術論文、研究報告書、学会による疫学調査・疾病登録、研究所・自治体によるサーベイランス・疾病登録に分類された。

疾患全般を包括的に対象としているものとして、人口動態調査（死亡率）、患者調査（受療率）、国民生活基礎調査（有訴率・通院者率）の3つの政府統計の他、学童期では学校保健統計調査などがあった。

研究実施の際には、先行研究の整理が不可欠である。成育保健・医療分野における各種調査研究の目録が継続的に維持・公開されることで、多くのリソースを用いて実施されてきた各テーマの研究成果を、一元的に把握することが可能となった。

2) レセプト傷病名の類型化に関する研究

医療情報の二次利用の重要性が昨今注目されている。診療情報明細書（レセプト）データベースや小児慢性特定疾病登録データベースと言った医療情報データベースは、データベースどうしのリンケージにより、その価値を高めることができると考えられるが、その際には疾患名によるリンケージができることが必須となる。レセプトデータベースは、同一の疾患概念であっても複数の傷病名により登録が可能となっているこ

とから、レセプトデータベースを二次利用するに当たっては、レセプト傷病名の類型化は避けて通れない。しかしながら、レセプトデータベースは極めて巨大であることから、従来の文字列検索のアプローチでは、膨大な時間と労力が必要となり、取扱が極めて困難な状況となっている。本研究は日本語に対応した自然言語解析の技術の一つである **InterSystems IRIS Natural Language Processing (NLP) Japanese** を利用してレセプト傷病名の自動類型化が可能であるかを検証した。NLP は主たる意味を持つ語句

を認識し、さらに語句欠損や入れ替えがあっても標準病名との対応を自動的に行った。NLP を利用した傷病名の類型化は、現実的な解法の一つとなり得ることが示された。

4. 研究内容の倫理面への配慮

目録作成については、基本的に公開ドメインに存在する各種データセットの系統的検索とそのプロファイルのみの利用であるため、特段の倫理的配慮を必要としない。レセプト解析については、国立成育医療研究センター倫理審査（1729）の承認を受けた。