

平成 28 年度厚生労働科学研究費補助金
(循環器疾患・糖尿病等生活習慣病対策総合研究事業)
分担研究報告書

日本の介護・医療政策における行政データの活用へ向けて
— 『介護給付費実態調査』・『人口動態調査(死亡票)』の Linkage(照合)の試み -

研究分担者 野口晴子 早稲田大学 政治経済学術院 教授
研究分担者 高橋秀人 国立保健医療科学院 統括研究官
研究分担者 川村顕 早稲田大学 政治経済学術院 准教授
研究協力者 富蓉 早稲田大学 政治経済学術院 助手
研究代表者 田宮菜奈子 筑波大学 医学医療系 教授

研究要旨

本研究では、「地域包括ケアシステム」の構築に資する日本の介護・医療政策における行政データの活用へ向けての試みの1つとして、『介護給付費実態調査』・『人口動態調査(死亡票)』等の行政データを Linkage(照合)する手法を確立し、今後の境界期健康寿命の分析への活用可能性を探ること目的とする。現在、厚生労働省・政策統括官(統計・情報政策担当)による『介護給付費実態調査』・『人口動態調査(死亡票)』の二次利用データを統計法第33条により申請中である。

A. 研究目的

地域包括ケアシステムの構築にとって、医療と介護の連携を図る手段としての ICT(information and communication technology)の活用は必須である。2017年1月9日、厚生労働省では、塩崎厚生労働大臣を本部長とする「データヘルス改革推進本部」を設置して、医療と介護のデータを一元化し健康管理の研究などに活用するための具体的な議論を開始することとなった。こうした動きは、当該課題に対する大きな前進といえるだろうが、そのための法的整備、及び、データ構築の手法については、今後の大きな課題である。

本研究では、こうした政策の動向を踏ま

え、日本の介護・医療政策における行政データの活用へ向けての試みの1つとして、『介護給付費実態調査』・『人口動態調査(死亡票)』等の行政データを Linkage(照合)する手法を確立し、今後の境界期健康寿命の分析への活用可能性を探ること目的とする。現在、厚生労働省・政策統括官(統計・情報政策担当)による『介護給付費実態調査』・『人口動態調査(死亡票)』の二次利用データを統計法第33条により申請中である。

B. 研究方法

データ照合の手法 - probabilistic linkage と deterministic linkage -

異なる目的で収集された行政データを照合するためには、照合されるデータファイルに、なにがしかの共通する情報が含まれている必要がある。相澤他（2004）によれば、「異なる情報源の間で共通するレコードを照合する問題は歴史が古く」、1940年代後半から1950年代にまでさかのぼることができる。ここでは、代表的なデータ照合の手法として、Fellegi and Sunter（1969）による probabilistic linkage と deterministic linkage の2つを挙げておこう。

probabilistic linkage では、異なる情報源のデータに含まれる各要素について、照合群（match）と非照合群（unmatched）を機械的に判定し、照合可能性あり群（possible match）については、アプリアリに閾値を設定し人手によって照合判定を行う。ここで、照合可能性ありは2つの確率から計算できる。1つ目は、本来異なる要素を偶然誤って同一要素と判定する確率、つまり、 $P(\text{照合} | \text{非照合})$ という条件付き確率として表現される。2つ目は、本来同一の要素を記録の登録ミスなどによって異なる要素と判定する確率、つまり、 $P(\text{非照合} | \text{照合})$ となる。Fellegi and Sunter モデルでは、これら2つの確率が定められた時、人手による判定を行う照合可能あり群に含まれる要素の数を最小にする戦略がとられる（相澤他，2004）。

probabilistic linkage については、米国疾病管理予防センターにおいて LinkPlus Beta version と呼ばれる照合ソフトウェアパッケージが開発されていて、無料でダウンロードが入手可能である（Thoburn and Yemane, 2015）。他方、deterministic linkage は、より単純に、異なる情報源のファイルに共通す

る、都道府県・市区町村・出生年月日・性などの識別変数群が存在する場合、全ての識別変数が合致すれば照合と判定し、1つでも合致しなければ非照合と判定する手法である。

Smith（1984）は、いずれの手法に妥当性があるかについて、次の3つの基準を提示した。1つは、データに含まれる識別変数群に間違いや欠損がどの程度含まれているか、第2に、識別変数に個人や組織を識別する能力がどの程度あるか、第3に、標本数がどの程度か、という基準である。Zhu et al.（2015）は、probabilistic linkage と deterministic linkage の2つの手法による照合を実施し、Smith（1984）が提示した3つの基準について感度分析を行い、データに含まれる識別変数群に間違いや欠損が少なく、個人を識別する能力が高く、さらに、標本数が多いビッグデータの場合は、deterministic linkage の方の妥当性が高いという結果を得た。他方、これら3つの条件が満たされていないければ、probabilistic linkage による照合の方が優れていると結論づけている。こうした感度分析は、個人を完全に識別可能な共通番号による照合が可能で、「真」の照合結果が分かっているなければ、行うことはできない。マイナンバーを利用することができない現段階では、日本でこうしたデータ照合の手法に関する比較研究の実施は極めて困難である。しかし、相澤他（2004）が指摘するように、「社会に氾濫する情報を整理し、役に立つ形で発信して行く」ためにも、また、たとえば、利用者を中心に据えた複合的な公共サービスの提供体制の実態をとらえるためにも、さまざまな情報を照合するシステムを構築す

る必要性は、今後、ますます高まるだろう。

(倫理面への配慮)

厚生労働省による『介護給付費実態調査』・『人口動態調査(死亡票)』の二次利用データを統計法第33条により申請し、許可を得て個票を分析した。提供された個票には個人を特定できる情報は含まれていない。本研究の実施にあたっては、「筑波大学医学医療系 医の倫理委員会」による承認を受けた。

C. 研究結果 / D. 考察 / E. 結論

現在、厚生労働省・政策統括官(統計・情報政策担当)による『介護給付費実態調査』・『人口動態調査(死亡票)』の二次利用データを統計法第33条により申請中である。研究結果と考察については、データを入手し次第としかかる予定である。

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

なし

H. 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

相澤彰子・高須淳宏・大山敬三・安達淳(2004)「レビュー・異種データベース間でのレコード照合に関する研究動向」. NII Journal No.8.

Fellegi IP, Sunter AB. (1969) "A theory for record linkage". *Journal of the American Statistical Association*, 64(328): pp.1183-1210.

Smith ME. (1984) "Record linkage: present status and methodology". *Journal of Clinical Computing*, 13: pp. 52-69

Thoburn, KK and Yemane S. (2015) "Section 3: Linkage Software and State Experiences". In: NAACCR Discharge Data Work Group (eds). *Discharge Data Best Practices Guide*; pp.11-12. Springfield, IL: North American Association of Central Cancer Registries.

Zhu Y, Matsuyama Y, Ohashi Y, and Setoguchi S. (2015) "When to conduct probabilistic linkage vs. deterministic linkage? A simulation study". *Journal of Biomedical Informatics*, 56; pp80-86.