平成27年度 厚生労働科学研究費補助金（成育疾患克服等次世代育成基盤研究事業）
「今後の小児慢性特定疾患治療研究事業のあり方に関する研究」
分担研究報告書

小児慢性疾患対策の検討及びデータの精度向上に関する研究

研究分担者 森 臨太郎（国立成育医療研究センター 政策科学研究部 部長）

研究要旨

都道府県格差を数値化することで、小児慢性特定疾患治療研究事業登録の都道府県格差が算出できる可能性を検討するため、人口動態統計による小児の死亡率の都道府県格差を算出し、社会背景的説明ができるかの検証を行った。死亡数、あるいは、関連死因による死亡の都道府県格差を、人口動態統計を用いて算出し、同じ手法を用いて小児慢性特定疾病における登録の格差などを算出することは理論的に可能であることが判明した。今後においては小児慢性特定疾病のデータを用いて算出し検証する。

さらに、完全一致によるレコードリンケージだけではなく、今後小児慢性特定疾患治療研究事業のデータをより高い確率で縦断データ化するため、確率的レコードリンケージを行うことで、悉皆性が高められるかどうかの検証を行った。この検証では、確率的レコードリンケージを用いることで、同じ変数を使いつつも、一致率が高まった。こういった手法を積み重なることで、小児慢性特定疾病の特性を踏まえた解析が可能になることが判明した。今後は、データの縦断化を目指し、さらに分析を進める。

研究協力者：

盛一 享徳（国立成育医療研究センター
臨床疫学部 研究員）
野間 久史（統計数理研究所データ科学研究系
助教）

A．研究目的

本研究は、今後小児慢性特定疾患治療研究事業にかかわるデータの利用を促進し、対策に資する研究が促進されるために、小児慢性特定疾患の登録データの精度向上を目的に、以下の二つの研究を行った。

研究 1)

都道府県格差を数値化することで、小児慢性特定疾病登録の都道府県格差が算出できる可能性を検討するため、人口動態統計による小児の死亡率の都道府県格差を算出し、社会背景的説明ができるかの検証を行った。

研究 2)

完全一致によるレコードリンケージだけではなく、今後小児慢性特定疾病のデータをより高い確率で縦断データ化するため、確率的レコードリンケージを行うことで、悉皆性が高められるかどうかの検証を行った。

B．研究方法

研究 1)

厚生労働統計協会が整理した人口動態統計における 1889 年より 2013 年までの都道府県別年齢別死亡数を用いて、わが国の 1889 年から 2013 年までの、0 歳未満死亡（deaths of infants）および、5 歳未満死亡（deaths of children undre 5 years）に concentration index を年ごとに算出し、年次推移を表示した。今回は男女の総数で算出した。両者とも総数のみで、出生都道府県不明のデータおよび海外出生・海外死亡のデータは除外した。

研究 2)

レコードリンケージソフトウェア（CDC Link Plus および Febrl）を用いて、小児慢性疾病登録データである、慢性腎疾患群の平成 23 年度および平成 24 年度登録データを利用し（H23 年度登録総数 8845 件、H24 年度登録総数 9008 件）、単独パラメータでは同一年度のデータセット内での重複値が多いことを、受給者番号および生年月日を用いて示し、確率的レコードリンケージ（Probablistic record linkage）によりアルゴリズムを作成し、実施主体番号＋受給者番号＋生年月日　の組み合わせデータでの一致率の向上を、検証した。

（倫理面への配慮）

研究 1) は、公開されているデータを用いた、二次的なデータ分析であり、特別な倫理的配慮は必要ないものと判断した。研究 2) は小児慢性特定疾病データを用いた理論的データ分析であり、同様に特別な倫理的配慮は不要と考えられた。

C．研究結果

研究 1)

まず 2013 年の都道府県別乳児死亡（1 歳未満児死亡）の分布が、出生数の都道府県別分布と比較して、どれくらいばらつきを認めるか、すなわち、死亡率の都道府県格差について、concentration index という経済学領域で確立された格差の評価指標を用いて算出した（図 1）。青色部分と赤色部分の間の面積が concentration index である。

その後、この concentration index について、1889 年から 2013 年まで、都道府県別乳児死亡数と幼児死亡数（5 歳未満児死亡）について算出し、推移を示した（図 2）。

Concentration index は、0 歳未満では年ごとのばらつきは少ないが、5 歳未満ではばらつき大きい。0 歳未満での傾向としては、戦前は少しずつ改善傾向を示していたが、戦後大きく悪化し、高度成長期に大幅に格差が是正され、バブル期に格差がほぼゼロになっていたが、2000 年以降急速に悪化をしてきている。5 歳未満児死亡においても同様の傾向が認められた。

研究 2)

A.単独パラメータでは同一年度のデータセット内での重複値が多いことの証明

①受給者番号のみでの重複調査

同一年度のデータセット内で 2 つ以上の重複（同一番号が出現）する件数は以下となった。

H23 年：4539 件（重複ペア数：6042 pairs）

H24 年：4895 件（重複ペア数：6689 paris）

②生年月日のみでの重複調査

同一年度のデータセット内で 2 つ以上の重複（同一の生年月日が出現）する件数は以下となった。

H23 年：重複ペア数 6599 pairs

H24 年：重複ペア数 6958 pairs

従って受給者番号や生年月日といった単独のパラメータで年度を越えたデータ結合は不可能であった。

B. 実施主体番号＋受給者番号＋生年月日 の組み合わせによるデータ結合（現在の小慢にお

ける同一患者同定ルーチン）

確率的手法を用いない古典的なレコードリンケージ手法を用いた場合、H23 年度データとH24 年度データを、都道府県＋受給者番号＋生年月日 の組み合わせで連結した場合、6758 件のデータが連結された。

次に下記の段階を経て、確率的レコードリンケージを、「実施主体番号＋受給者番号＋生年月日」の組み合わせデータで行った。

①受給者番号のみ異なり、他のパラメータは完全一致（実施主体番号、保健所番号、生年月日、性別、登録病名）している場合は、受給者番号の再発行と思われた。

②生年月日データのうち、年、月、日のいずれか一つのみ異なり、他のパラメータは完全一致（受給者番号、実施主体番号、保健所番号、性別、登録病名）している場合は、ヒューマンエラーによる入力ミスと思われた。

以上のケースで、各々の独立フィールドから得られるスコア値の合計が一定の範囲内に収まる場合を matched data と判断し、最終的に6919 件のデータが連結された。すなわち、Probablistic linkage の手法により、6758 件から6919 件へ 約 2.4% 一致件数を向上させることが可能であった。

D．考察と結論

研究 1)

死亡数、あるいは、関連死因による死亡の都道府県格差を人口動態統計を用いて算出し、同じ手法を用いて、小児慢性特定疾患治療研究事業登録データにおける登録の格差などを算出することは理論的に可能であることが判明した。さらに、今後においては、登録データを用いて算出し検証することとしたい。

研究 2)

確率的レコードリンケージを用いることで、同じ変数を使いつつも、一致率が高まった。こういった手法を積み重なることで、小児慢性特定疾患治療研究事業の特性を踏まえた解析が可能になることが判明した。今後は、さらにデータの縦断化を目指し、分析を進めてゆきたい。

E．健康危険情報

なし

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

1. 特許取得／2. 実用新案登録／3.その他
   いずれもなし

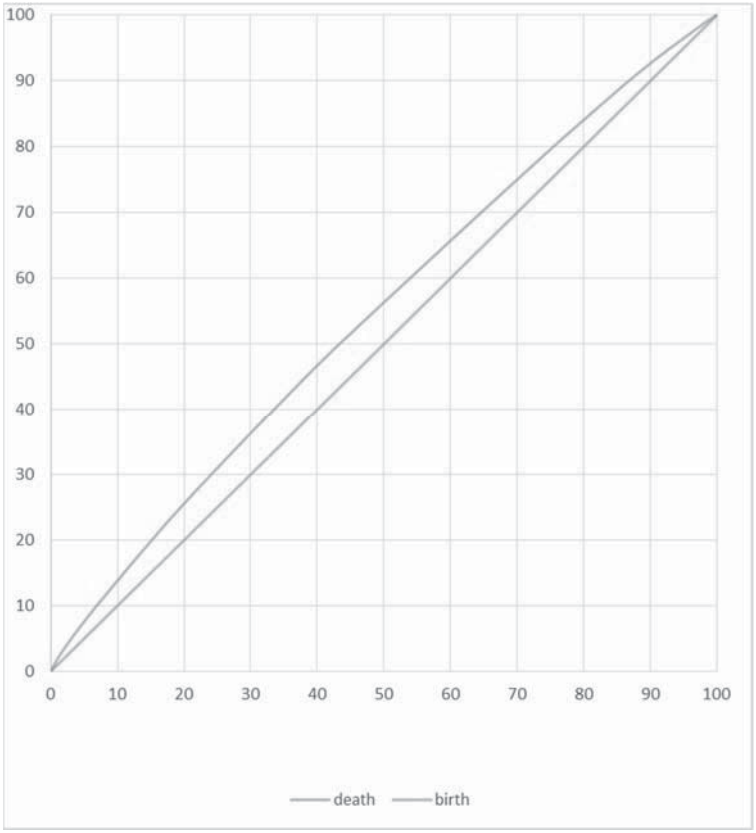

図 1 死亡率の都道府県格差についての concentration curve

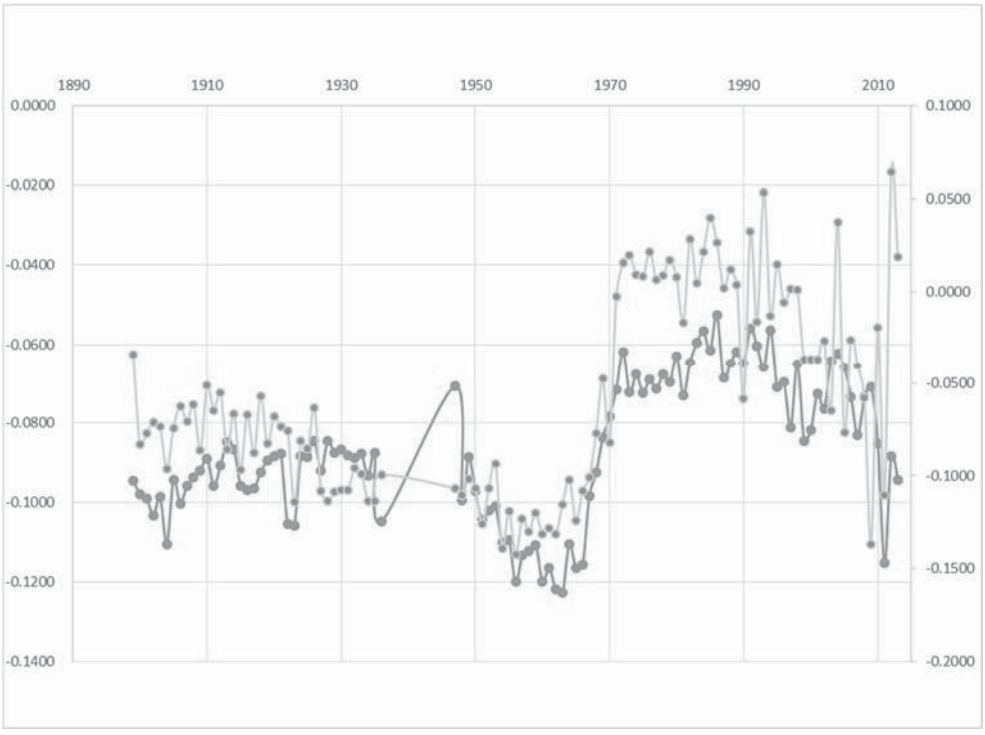

図 2　1889 年から 2013 年までの都道府県別、乳児死亡数および幼児死亡数（5 歳未満児死亡）に関する Concentration Index の変化