

平成 26 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 (健やか次世代育成総合研究事業))
「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

小児慢性特定疾患データの質向上のための個人連結番号に関する研究

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 データを利活用してより精緻な政策や研究、診療に資する科学的根拠づくりに重要である。そのためには、「個人を特定せず、連結を可能とする」変数を各データベースが保持しておくことが重要である。そこで、以下を条件として、出生届のデータを用いて、上記条件に合致する最も適切な個人連結変数の組み合わせを算出した。1) 母子健康手帳に記載されているなど、日常臨床の現場で情報の取得が容易であること、2) できるだけ少ない変数の組み合わせで、個人が連結できること、3) 成長・発達や社会的環境が変化しても (別の言葉で言えば一生を通して)、変化しないこと、4) 連結はできるが個人を特定しないこと。以上により、「出生都道府県・児の生年月日・出生体重・性別・母の出産時年齢 (年)・在胎週数」を共通項目があれば、研究利用のためという条件下で個人を特定することなく連結することができると考えられた。

研究協力者:

森崎 菜穂 (国立成育医療研究センター社会
医学研究部 ライフコース疫学
研究室長)

盛一 享徳 (国立成育医療研究センター政策
科学研究部 研究員)

A. 研究目的

データを利活用してより精緻な政策や研究、診療に資する科学的根拠づくりに重要である。そのためには、「個人を特定せず、連結を可能とする」変数を各データベースが保持しておくことが重要であり、医療意見書の共通項目として、適切な共通個人連結変数を検討することを目的とした。

B. 研究方法

小児慢性特定疾患のデータベースは悉皆性やデータの信頼性において改善の余地があるとされている。しかしながらそれぞれの特性

を持つデータは存在していても、すべての側面において強い特性を持つデータは存在しない。

異なるデータベース間で共通する変数を突合せ、個人を特定することなく連結する手法 (レコードリンケージ) の利活用が進んでいる。リンケージ後の妥当性評価や右図にあるように推測モデルを用いることで、それぞれのデータセットの特性を最大限に生かし、悉皆性や信頼性の高いデータをもとにした分析が可能となる。

小児慢性特定疾患データベースも、今後、出生届・死亡届・乳幼児栄養調査などの政府統計、各関連学会が行っている疾病登録などとの連結を可能としておくことが、データを利活用してより精緻な政策や研究、診療に資する科学的根拠づくりに重要である。そのためには、「個人を特定せず、連結を可能とする」変数を各データベースが保持しておくことが重要であり、医療意見書の共通項目として、共通個人連結変数を含めることを検討した。

そこで、共通個人連結変数の条件としては、

- 1) 母子健康手帳に記載されているなど、日常臨床の現場で情報の取得が容易であること
- 2) できるだけ少ない変数の組み合わせで、個人が連結できること
- 3) 成長・発達や社会的環境が変化しても（別の言葉で言えば一生を通して）、変化しないこと
- 4) 連結はできるが個人を特定しないことを条件とした。

なお、本文脈においては、個人を 100%の一致率で特定することが義務付けられる「行政的」な連結変数ではなく、100%は一致しなくとも、全般的な分析として一定レベルの精緻さを確保する「研究的」な連結変数が適切と考えられる。

出生届のデータを用いて、上記条件に合致する最も適切な個人連結変数の組み合わせを算出した。

（倫理面への配慮）

本研究は観察研究にあたるため疫学研究の倫理指針に即し、国立成育医療研究センター倫理委員会において倫理審査の承認を得て行った。

表:各変数ごとのリンケージ成績

変数	単独最大重複数	全項目重複数	重複あり人数（欠損値なしのみ、100万人中）
出生都道府県	108377 (市町村を含めると2000)	全8項目のデータで重複するのは最大3名で、これが600人（100万人中）となる	15000
児の生年月日	4083		必須と考えられる
出生体重	11441		必須と考えられる

C. 研究結果

まず、母子健康手帳や出生届にある変数で、条件3)・4)に合致し、2)の可能性が高いものは出生地・児の生年月日・出生体重・性別・母の出産時年齢・多胎の場合に出生順位・在胎期間となる。これらについて、最も悉皆性の高い出生届1年分のデータを用いて検討を行った。（表参照）

D. 結論

以上の検討より、出生都道府県・児の生年月日・出生体重・性別・母の出産時年齢（年）・在胎週数を含む共通項目があれば、研究利用のためという条件下で個人を特定することなく連結することができると考えられた。

E. 健康危険情報

F. 研究発表

G. 知的財産権の出願・登録状況（予定も含む）

なし

性別	552711	る。在胎日数を含めると重複は無くなる。市町村を入れたら重複は最大2名で、これが10名となる。	必須と考えられる
母の出産時年齢	80358		12000
双胎以上の場合は出生順位	10104		690
在胎週数	298211 (日数を含めると2000)		2600