

厚生労働科学研究費補助金
政策科学総合研究事業（臨床研究等 ICT 基盤構築・人工知能実装研究事業）
分担研究報告書

各種厚生労働省統計と周産期関連学会データベースのリンケージと解析

分担研究者 森崎菜穂 国立成育医療研究センター社会医学研究部 室長
研究協力者 大久保祐輔 国立成育医療研究センター社会医学研究部 共同研究員
山本 依志子 国立成育医療研究センター政策科学研究部 上級研究員
小川浩平 国立成育医療研究センター産科 医員

研究要旨

本分担研究においては、人口動態統計の出生票、死産票、および死亡票をリンケージする複数の手法を比較検討することで、もっとも正確にこれらをリンケージできる手法を提案すること、そして、各種の周産期関連データベースをリンケージしたデータベースの利用を促進し、その解析を通して単一のデータベースからは産出不可能であった医学的なエビデンスを複数提示すること、を目的としている。

2年目である本年度は、2013-2015年度に出生した児の出生票と母の死亡票を高精度にリンケージするための方法を検討し、2003-2011年度の出生児について日本産科婦人科学会周産期登録データベース、新生児医療ネットワーク登録データベース、出生票、死産票、乳児死亡票を連結したデータベースを様々な角度から解析し、妊婦および児の予後に関係する医学的・社会的因子について、複数のエビデンスを発表した。

A . 研究目的

本分担班では、昨年度は人口動態統計の出生票と乳児死亡票を連結する手段を模索した。しかし、その際、氏名情報などの特異度が高い情報が欠落しているデータベース同士では、連結の不確実性がどうしても高いことが明らかになった。

出生、死亡、などの人口動態調査については、自治体から厚生労働省に送付される調査票のOCRシートを読み込む方法のほかに、自治体において調査票の電子データをオンラインで送信する人口動態統計オンライン報告システムが平成15年より導入され、各市町村に順次導入されている。平成

26年度時点において、全出生の約90%が本報告システムを用いて報告されている。

そこで、本年度の本分担研究班における研究目的は、出生票とその母の死亡票を高精度にリンケージする手法を提案すること、各種の周産期関連データベースをリンケージしたデータベースの解析を通して複数の医学的に有用なエビデンスを提示すること、である。

B . 研究方法

出生票とその母の死亡票を高精度にリンケージする手法の検討

現在、出産後の女性の死亡が妊娠に関係

する死であったかどうかを判定するための材料は、死亡票に記載されている死因情報のみに依存している。しかし、この方法だけでは、出産後時間の経った症例では特に見落としが起きる可能性がある。

このため、妊婦死亡の把握（永田班）および妊婦自殺の把握（大田班）の解析データセットを作成するために、妊娠可能年齢の女性の死亡票・個票を、その妊娠の結果出生あるいは死産となった児の出生票・出生個票あるいは死産票・死産個票と連結するための手法の検討を行った。

2003-2011 年度出生において、日本産科婦人科学会周産期登録データベース、新生児医療ネットワーク登録データベース、出生票、死産票、乳児死亡票を連結したデータベースを、複数の研究者で解析した。

（倫理面への配慮）

本研究は二次的に得られる情報で行う研究であり、情報収集については特別の倫理的配慮は必要としなかった。しかし、個人情報を含む情報の解析であるため、国立成育医療研究センターの倫理委員会において研究計画の承認を得た後に行い、情報漏えいリスクを最小限にとどめるために外部ネットワークから遮断された環境において解析を行い、また結果公表に際しても5例以下のセルについては報告を行わないことで少数例庇護の措置を行った。

C . 研究結果

出生票とその母の死亡票を高精度にリンケージする手法の検討

現在、日本の人口動態統計は市町村が厚生労働省に各種出生・死亡などの情報を個

票という形で報告し、これを厚生労働省がクリーニングしたのち、氏名情報や病院名情報を削除し、死因コードを ICD 分類に基づき付与した結果のデータを公式統計に用いている。また、市町村から厚生労働省に提出される情報は約 9 割がオンライン報告となっている。

このため、個票のみに氏名情報や死因の詳細情報が載っており、その代わり、誤った情報が記載されていることも多く、また日本全体の 10%の症例については情報がないという状態にある。親子のリンケージを行うためには死亡個票や出生個票のみに含まれる母親の氏名情報も、死亡票のみに記載されている死因の ICD 分類も必要であるため、本リンケージは 2 段階で実施した。

統計法（平成 19 年法律第 53 号）第 33 条の規定に基づき、人口動態調査に係る調査票情報の提供について申出を行い、調査票情報の提供を受けた。

まずは市町村からオンラインで報告されるデータ（個票）およびこれを厚生労働省がクリーニングおよびコーディングした結果のデータ（票）を、出生、死産、死亡のそれぞれにおいて作成した。

このリンケージには、届け出のあった都道府県・市町村・保健所・支所番号・事件簿番号および（出生・死亡・死産の）年月日、を用いて、年月日および事件簿番号に書き間違いがあった可能性を 1% と仮定して、確率的リンケージ(Probabilistic Linkage)により連結することで作成した。確率的リンケージについては Fellegi and Sunter らによって提唱された理論を用いて、誤字や入力ミスについては編集距離(Lebenstein 距離)を用いて 2 値の一致度を評価した。

下記の図 1-図 3 に出生、死亡、死産の

それぞれについて、リンケージ結果を示す。なお、m-probability (変数に書き間違いがない確率)を 99%から上下 2%変化させても、リンケージされる確率はほとんど変わらなかった。

図 1 出生票と出生個票のリンケージ

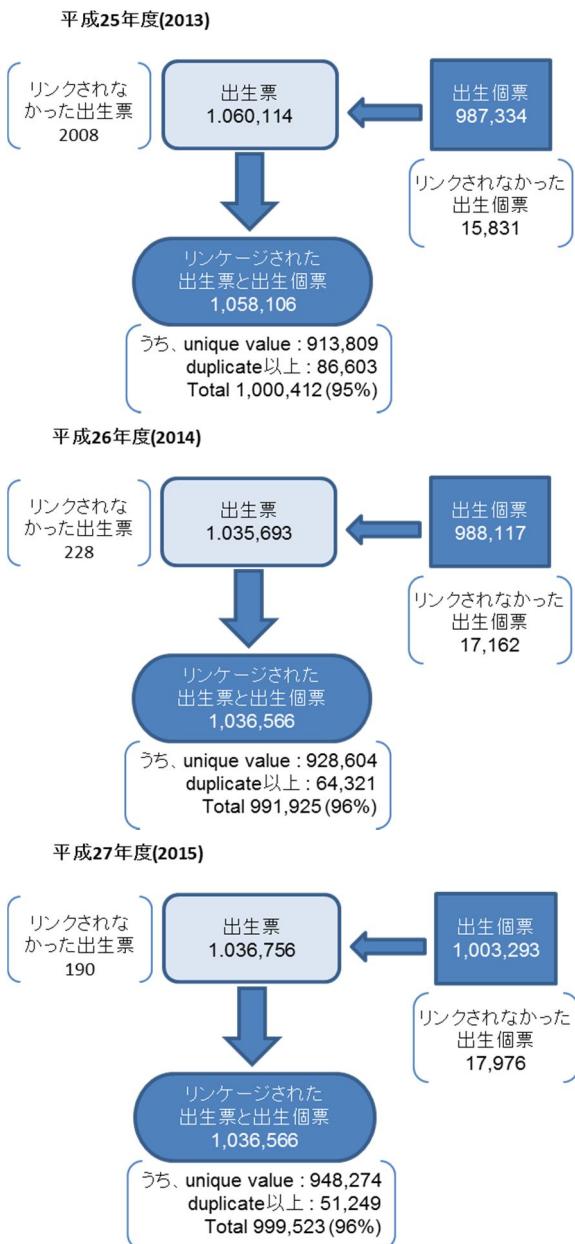


図 2 死亡票と死亡個票のリンケージ

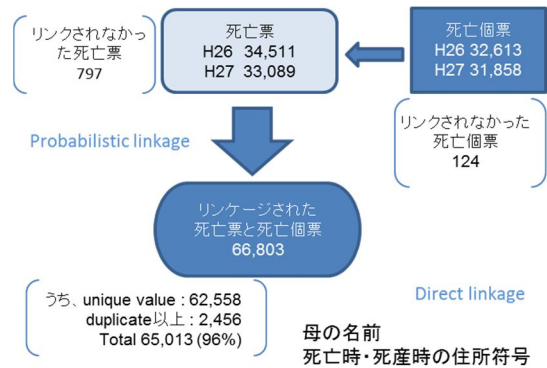
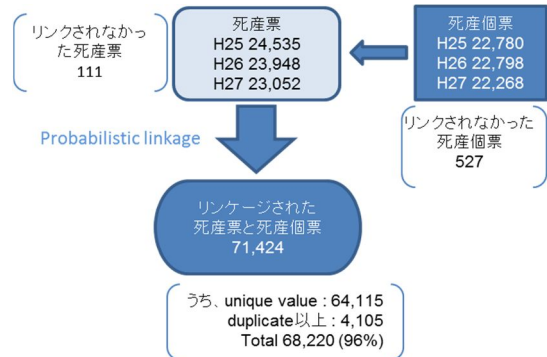


図 2 死亡票と死亡個票のリンケージ



続いて、それぞれリンケージされた死亡票/死亡個票と出生票/出生個票を、母(女性)の氏名および生年月日によりリンケージした。また、それぞれリンケージされた死亡票/死亡個票と死産/死産個票を、母(女性)の氏名および死産時の年齢(+1 あるいは-1 を含める)によりリンケージした。

このリンケージには、Deterministic Linkage つまりは、完全一致していることを条件とした。この方法では結婚あるいは離婚により出産時と死亡時で母の苗字が変わっている場合は捉えられない。しかし、下の名前のみを頼りに連結を試みると、たとえ同じ市町村内での出産・死亡であるという制限をかけたとしても下

の名前も生年月日も同じであるという女性が複数いたため、この方法を用いてのリンケージは現実的ではなかった。

リンケージされたデータの利活用

本分担研究関係者のみならず、他の分担研究の先生方とともに多角的な解析を積極的に行った。この結果、複数の英語原著論文を出版することが出来、現在も複数を投稿中である。

今年度は論文および学会発表にて、下記の事項を報告した。

-低身長妊婦が、より身長が高い妊婦よりも早産のリスクが高い理由は、身長が低いほど妊娠高血圧腎症になるリスクが高いことにより説明される。

-日本人の平均身長は1979年を境に減少しており、生まれ年別に見ると平均身長はその年の低出生体重児率と逆相関を認める。

-高齢出産は早産や多くの妊娠合併症のリスクとなる。

-日本を含む先進国34カ国における早産率と37-38週の出生率は高い相関率を示しており、37-38週での出生を抑えることが早産率減少にも役立つ可能性がある。

-22-24週の分娩において、児が分娩中の死産あるいは1時間以内の新生児死亡となるリスクは、母の居住市町村の平均年齢が低いほど高い。

さらに、なるべく幅広くこのデータベースを有効に活用していただけるように、周産期医療関係者への疫学教育を実施した。

D. 考察・結論

本年度は人口動態統計の個票情報に

含まれる氏名情報などの特異度の高い情報を用いることで、それまでこれらの情報が含まれていない出生票・死亡票のみでは不可能であった親子を連結する手段を検討し、これが有用である可能性を見出した。

また、事件簿番号が不一致している症例が一定の割合でいる可能性があることやまだ個票のオンライン報告率が100%ではないなどのいくつかの問題がみつき、今後はその不一致の理由の探求などが必要であることがわかった。

今までは人口動態統計の連結手段に関しては、匿名化されている人動態統計票の連結が必要であった。これらの情報を高精度に相互連結するためには、母の生年月日や、周産期関連因子など、現在1歳未満の死亡の特記事項として記載されている変数が必要であり、これらの変数がないと出生票と死亡票の正確な連結は難しかった。

今回、氏名情報を用いた家族の連結が可能であることが分かった。この方法を用いて、出産や中絶後の母の死亡のリスク因子の解明に役立つ可能性が高いことが示された。今後兄弟の同定も可能であると思われる、家族のリンケージが進めば、幼児、学童の死亡についても同様の検討が可能になると思われる。

一方で、個人を特定できる可能性が高く“個人識別符号”とされている氏名情報を用いても、結婚や離婚により姓が変わりうる妊婦や褥婦においては、異なるデータベースに含まれる同一人物を完全に特定するのは難しい可能性も示唆された。現在戸籍情報の登録にはマイナンバーは含まれていないが、将来的にはマイナンバーによ

り戸籍情報を含む政府所有の情報を高精度で連結し、公衆衛生学的に有用な研究に活かすことを可能とすることが必要かもしれない。

連結されたデータベースの利活用については、本年度も小児科および産科の先生方とともに周産期データベースの解析を行い、多くの有用な新規発見ができた。しかし、有用なデータベースが作成できても、それを活用し、臨床現場および政策に反映できるような成果を産出できる研究者はまだ少ないという問題もまだ依然として残っている。

E. 健康危険情報

該当なし

F. 研究発表

1. 論文発表

- 1) Morisaki N, Urayama KY, Yoshii K, Subramanian SV, Yokoya S. Ecological analysis of secular trends in low birth weight births and adult height in Japan. *J Epidemiology and Community Health*. 2017 Oct;71(10):1014-1018.
- 2) Morisaki N, Ogawa K, Urayama KY, Sago H, Sato S, Saito S. Preeclampsia mediates the association between shorter height and increased risk of preterm delivery. *International Journal of Epidemiology* 2017 Oct 1;46(5):1690-1698.
- 3) Ogawa K, Urayama KY, Tanigaki S, Sago H, Sato S, Saito S, Morisaki N. Association between very advanced maternal age and adverse pregnant outcomes : a cross sectional Japanese

study. *BMC Pregnancy and Childbirth* 2017 Oct 10;17(1):349

4) Helenius K, Sjors G, Shah PS, Modi N, Reichman B, Morisaki N, Kusuda S, Lui K, Darlow B, Bassler D, Hakansson S, Adams M, Vento M, Rusconi F, Isayama T, Lehtonen L. Survival in very preterm infants: an international comparison of 10 national neonatal networks. *Pediatrics*. 2017 Dec; 140(6). pii: e20171264.

5) Delnord M, Mortensen L, Hindori-Mohangoo A, Blondel B, Gissler M, Kramer MR, Richards JL, Deb-Rinker P, Morisaki N, Nassar N, Nybo Andersen AM, Kramer MS, Zeitlin J. Can we apply a population approach to preterm birth prevention? An ecological study of preterm and early term births in 34 high-income countries. *European Journal of Public Health*. 2018 Apr 1;28(2):303-309.

6) Morisaki N, Isayama T, Samura O, Wada K, Kusuda S. Socioeconomic inequity in survival for deliveries at 22-24 weeks of gestation. *Archives of Diseases in Childhood. Fetal and Neonatal Edition* 2017 Aug 28. doi: 10.1136/archdischild-2017-312635. [Epub ahead of print]

7) Yamaoka Y, Morisaki N, Noguchi H, Takahashi H, Tamiya N. Comprehensive assessment of risk factors of cause-specific infant deaths in Japan. *Journal of Epidemiology* Feb 10, 2018 doi: 10.2188/jea.JE20160188 [Epub ahead of print]

2 . 学会発表

- 1) Morisaki N, Isayama T, Samura O, Wada K, Kusuda S. Socioeconomic inequity in fetal and infant survival at 22 to 24 weeks of gestation. Pediatric Academic Societies 2017 Meeting, May 6 - 9, San Francisco, CA
- 2) Morisaki N, Smith L, Morken N, Gissler M, Deb-Rinker P, Rouleau J, Hakansson S, Kramer MR, Kramer MS. Assessing the impact on reported survival rates of international variation in the classification of deaths at 22 to 26

weeks gestational age (GA) Society for Pediatric and Perinatal Epidemiologic Research 2017 Meeting, June 19 - 20, Seattle, WA

G. 知的財産権の出願・登録状況

1. 特許取得

該当なし。

2. 実用新案登録

該当なし。

3. その他

該当なし。