

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）

平成30年度（総合）統括研究報告書

周産期関連の医療データベースのリンケージの研究

研究代表者 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
(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
掛江直子 国立成育医療研究センター生命倫理研究室・室長
康永秀生 東京大学大学院医学系研究科公共健康医学専攻臨床疫学・教授
大田えりか 聖路加国際大学大学院看護学研究科・教授
永田知映 国立成育医療研究センター臨床研究開発センター -
臨床研究教育部・室長
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
Mahbub Latif 聖路加国際大学大学院 公衆衛生学教室・教授

本研究では成育医療分野における各種統計や医学団体所有データベースを過去に活用し、また周産期医療関係の各種データベースをリンケージする手法に関する研究を行ってきたという経験を生かして、成育医療分野のデータベースを連結することで拡充し、さらに多くの臨床研究に活用する。また、公的統計の妥当性検証やデータベース同士の連結手法を確立することで今後の研究基盤を作成することが目的であった。

まず、解析班では研究初年度に Probabilistic Linkage Method を用いて個票レベルでのリンケージに成功して作成した、2004-2011 年の出生票に乳児死亡票および周産期の全国データベース（日本産婦人科学会周産期データベース、新生児臨床研究ネットワークデータベース）に更に市町村単位での経済指標のデータベースを連結し、研究2-3年目に解析した。この大規模データベースを用いて、周産期予後に関連する因子に関する英語原著論文を複数発表した。データベースのリンケージが行われたことにより、これまで大規模の解析が行われていなかった研究を行うことができ、国際誌でも高く評価される結果となった。特に、低出生体重児増加による日本の平均身長推移の算出などの疫学研究結果は有名科学紙 Science 紙にも取り上げられるなど、科学界やメディアでも大きく取り上げられ、いずれも今後の政策の基礎資料となると思われる。

また、リンケージ・検証班では DPC データ、小児慢性特定疾病登録データの人口動態統計・臨床

レジストリとの個別リンケージの可能性評価を行い、DPC 情報による周産期医療に関するレジストリ情報の代替可能性を検討するなどリンケージを行うデータベースの更なる拡張を行った。人口動態統計については出生票・死亡票・死産票を高精度でリンケージする手法を複数比較し、研究 2-3 年目に実践することで“出産・流産 1 年以内に死亡した女性”を同定することに初めて成功した。本リンケージ結果を活用した“産後一年以内の死亡”の死因別把握はテレビ報道や各種新聞紙の 1 面を飾るなどメディア等でも大きく取り上げられた。

更に、これらのデータとのリンケージ体制の構築と実務および医療情報の二次利用による臨床研究を行う際に留意すべき現行法令や指針とまとめることを通して、医療情報のデータリンケージとその利活用の可能性と問題点とまとめた。本取り組みはデータリンケージを行っていく際の倫理的・法律的な問題への対応方法の参考となると思われる。

A. 研究目的

本研究では成育医療分野における各種統計や医学団体所有データベースを過去に活用し、また周産期医療関係の各種データベースをリンケージする手法に関する研究を行ってきたという経験を生かして、成育医療分野のデータベースを連結することで拡充し、さらに多くの臨床研究に活用する。また、公的統計の妥当性検証やデータベース同士の自動連結手法を確立することで今後の研究基盤を作成することが目的である。

これにより、医療計画の「5 疾病 5 事業」に含まれ、その医療体制整備は極めて重要である小児医療および周産期医療において、関連学会が積極的に作成してきたレジストリや政府が行ってきた政府統計をいまよりも更に有用に活用する方法が広がると考えられる。

B. 研究方法

研究 3 年間を通して、人口動態統計のリンケージの手法論の検討と実践を継続して実施されたが（主に永田・森崎班が担当）、生物統計専門家の協力が必要と判断し、研究 3 年目より該当分担研究者が増員された（ラティフ班）。DPC 情報のリンケージ可能性と妥当性評価は 2016-2017 年度に実施されたが（主に康永班が担当）当該研究班が用いている DPC 情報は他の臨床情報とのリンケージが不可能であることより、研究 3 年目には他分担研究班（森崎班）と合同で新生児関連臨床レジストリのデータリンケージ体制の考案を行った。また小児死因統計の臨床的死因との合致性に影響する要因に関する分担研究は、死亡票に記載された死因の正確性について研究 1-2 年目に解析を行ったが（溝口班）、本研究テーマを主題とする別研究班（厚生労働科学研究費補助金（健やか次世代育成総合研究事業）「突然の説明困難な小児死亡事例に関する登録・検証システムの確立に向けた実現可能性に

関する研究」)の立ち上がりにより、こちらの研究班で継続実施される形となり、本分担当研究(溝口班)は最終年度途中で削除された。人口動態統計と臨床レジストリ情報をリンケージしたデータベースの解析は3年間を通じて実施され、多くの臨床疫学論文が作成された。また、人口動態統計からの産後1年以内の死亡数および死因、自殺者の社会背景を算出は、人口動態統計内のリンケージ手法が確立された研究2-3年目に実施された(主に永田・大田班が担当)。法律面・倫理面では、3年間を通じて個別研究で用いられたリンケージ手法について、その倫理的妥当性の監督に加え、医療データベースのリンケージ研究実施に際して適用される法令等に関する検討が行われた(掛江班)。なお、旧研究代表者の職場異動にあたり、3年目の年度途中で研究代表者の交代が行われた。

C. 研究結果

小児慢性特定疾患データベースのリンケージと解析に関する研究

(分担: 森臨太郎)

本研究は、各種データのリンケージ及び利活用の推進にあたり、主に小児慢性特定疾患登録データベースと他のデータベースとのリンケージに関する検討を行った。

平成28年度は、小児慢性特定疾患レジストリを縦断的に連結し、さらに他のデータベースと連結するための分析を行った。平成29年度は、相対的格差指標であるTheil indexを用いて、小児慢性特定疾患

のデータ登録状況に地域間格差がないかを検討し、全般的に登録格差が少なく、慢性的に経過する内科的疾患ではとくに登録格差が大きくないことを示した。平成30年度は、小児慢性特定疾患登録データベースとレセプトデータベースとのリンケージを想定するにあたり、疾患名による関連付けが必須となることから、レセプト傷病名の機械的な類型化の可能性について検証を行い、自然言語解析の技術が解法の一つとなり得ることを示した。

DPC データベースのリンケージと解析

(分担: 康永秀生)

周産期関連の医療データベースのリンケージ研究を推進するために、H28年度は、匿名化情報のみで2つのデータベースのリンケージがどれくらい正確に行えるかを検討し、H29年度は、周産期関連データベース研究の重症度補正の際に重要となってくる小児慢性疾患診断名のDPCデータにおける代替性について検証した。

DPC データベースにおける年齢別周産期関連項目の入力率と、周産期関連データベースとのリンケージの可能性についての検討

【目的】単施設の医療情報を用い、周産期関連データが実際にどれくらい入力されているか、DPCデータとその他の周産期関連データとをどれくらい正確にリンケージできるか検証した。【方法】3年間のDPC情報から周産期関連項目の入力割合について年齢カテゴリごとに算出した。匿名化情報

のみで DPC 情報と患者基本情報、分娩情報とのリンケージを行い、患者 ID でリンケージした場合と比較し、感度・特異度を算出した。【結果】性別、身長、退院時転機は全ての入院情報で記録されていた。出生体重は、退院時年齢を 1 か月未満に絞るとは 98% で入力されていた。DPC 情報と患者基本情報は、感度 69.1%、特異度 79.3% でリンケージできた。DPC 情報と分娩情報は、感度 67%、特異度 100% でリンケージできた。【結語】DPC 情報における周産期情報の入力割合は非常に高く、他の匿名化された周産期データベースとのリンケージの実現可能性が高いことが示された。

DPC データにおける Pediatric complex chronic conditions classification system version 2 の妥当性検証

【目的】Pediatric complex chronic conditions classification system version 2 (以下 CCC) は、診断 (ICD-10) コードと、Procedure コードを組み合わせ、小児の慢性疾患を 12 に分類する手法である。単施設の医療情報を用い DPC データにおける CCC の妥当性を検証した。【方法】約 6 年間の死亡退院患者を電子カルテを用いて 12 の慢性疾患の有無について調べた。このデータをゴールドスタンダードとして、12 慢性疾患の有無に関して、CCC 自動分類の感度特異度を算出した。【結果】新生児疾患以外の慢性疾患の特異度は 90% 以上で非常に高かった。感度は、悪性腫瘍、新生児疾患、代謝疾患で高

かった。消化器疾患、神経疾患では感度は低いながら 60% 近くは保たれていた。一方デバイス依存の感度は 12% と非常に低かった。

【考察】移植患者の分類は現状の CCC 自動分類では分類できないことが分かった。

【結語】CCC 自動分類は DPC データ上でも非常に高い特異度と比較的高い感度を持って小児の慢性疾患を分類できることが示された。移植患者の分類に関しては現状の CCC 自動分類プログラムには問題があり、修正が必要であることが分かった。

周産期臨床データベースと DPC データを用いた、産科合併症に関する研究

(分担：永田知映)

本分担研究では、人口動態調査データを用いて妊娠中から周産期、小児期の健康や死亡について調査研究することを目的として、我が国における 5 歳未満死亡率の都道府県間格差の推移 (1899 年 ~ 2014 年) の検討と、人口動態調査 (出生票・死亡票・死産票) のリンケージによる母体死因に関する検討を行うこととした。

5 歳未満死亡率の都道府県格差の年次推移は、人口動態調査が始められた 1899 年から 2014 年までのデータについて、各都道府県の年毎の 5 歳未満死亡率を計算し、さらに格差を測る指標の一つである Theil index を年毎に計算した。5 歳未満死亡率の Theil index は第二次世界大戦後に上昇したのち徐々に下降して 1970 年代には 0.01 未満まで低下した。しかしながら 2000 年代に入って再び上昇しはじめ

2014年には1970年の値を超え、第二次世界大戦以前の値に近くなった。本研究により、子どもの健康において格差が拡大している可能性が示唆され、その原因、メカニズム、そして解決策に関する今後の研究が求められる。

人口動態調査データのリンケージによる母体死因に関する検討については、生殖可能年齢の女性の死亡票・死亡個票のリンケージにより死亡データベースを作成し、同様にして作成した出生データベースと死産データベースを死亡データベースとリンケージすることで抽出された症例、ICD-10コードや妊娠関連語句を用いて抽出した症例から、我が国の妊娠中から産後1年未満の女性の死亡の全体像を把握した。2015-2016年で357例の死亡例が見つかり、そのうち最も多かった死因は自殺であった。死亡データベースと出生・死産データベースのリンケージは、既存の制度や仕組みで把握が難しかった産褥婦の自殺例や、後期妊産婦死亡例の把握に有用な手段であった。一方で、妊娠中の死亡例については抽出できない、死亡診断書の記載のみでは死因の同定が困難である場合があるなどの限界も見出された。2017年から死亡診断書に妊婦または産後1年未満の産婦が死亡した場合は、産科的原因によるかを問わず、妊娠または分娩の事実を記載するよう改正されており、妊産婦死亡症例の把握率上昇が期待されている。今後、年次推移を追うことで、母子保健の指標の一つとして活用されることが期待される。

産褥婦の自殺にかかる状況及び社会的背景に関する研究

(分担：大田えりか)

妊産婦死亡のデータは、妊娠・出産に関連した原因によるものと定義されており、出産後、うつ病の悪化等により自殺に至った死亡は、わが国ではこれまで含まれておらず、これらの全国的な症例数は把握されていない。人口動態統計出生票及び死亡票の連結により抽出された、2015～2016年における生児出産後1年未満の褥婦の自殺例92件を抽出し、背景や自殺方法などを分析した。35歳以上、初産、及び世帯の職業が無職の女性において、最も自殺率が高かった。自殺の時期は、産後1年を通して自殺がみられた。人口動態調査出生票及び死亡票のリンケージは、産褥婦の自殺死亡例の把握に有用な手段と考えられる一方で、死亡診断書に記載される事項は限られているため、産褥婦の自殺死亡を予防するための対策に結びつけるためには情報が不十分であり、各症例についてさらなる詳細な調査が必要と考えられる。

小児死因統計の臨床的死因との合致性に影響する要因に関する研究

(分担：溝口史剛)

小児における死因統計上の死因と、臨床的な死因との合致性に関しての検討を行うため、初年度である平成28年度は既にある既存のデータ、具体的には、東京都・群馬県・京都府・北九州市を対象として実施し

た 2011 年の 15 歳未満の死亡事例（うち東京都は 5 歳未満事例）の後方視的検証（パイロットスタディー）の際に収集した情報のうち、乳児死亡事例 214 例のデータと、人口動態調査における乳児死亡単純分類統計上の死因との合致性につき調査を行い、死因事態が変更すべき事例（レッド事例）は 214 例中 58 例（27%）存在し、死因変更を要さないものの、欄や欄への追記を含む何らかの修正が望まれる事例（イエロー事例）も 214 事例のうち 48 例（22%）存在していたとの報告を行った。この結果を受け、死後に包括的な情報を集約したうえで、全年齢全数の死因検証を行う体制（チャイルド・デス・レビュー：CDR）の整備が不可欠であるとの考察を行った。

2 年目である平成 29 年度には、本年度は CDR を実施する上で、現在の各種法制度の下で収集された既存情報をどのように活用できるのかにつき検討した。現行では死亡小票の二次利用は困難であるが用いることで全数把握は可能であるが、その内容をもとにした要詳細検討事例のスクリーニングは不可能で、既存情報を生かすためには、別の法令根拠が求められると考察された。また現行法の弾力的運用では、関係法規とのバッティング（刑事訴訟法、個人情報保護法など）が生じるため、「チャイルドデスレビュー」という文言そのものを法令に記載し、根拠を明確にし、既存情報活用（共有）・新規情報の収集を可能とする必要があると思われた。このような情報のリンケージを進めるために、リンケージす

べき情報とその利活用の範囲につき明確化する必要があると考察した。

このように 1・2 年目の研究を通じ、小児死亡の情報リンケージのためには CDR が不可欠であることが明確化し、また同時期に CDR に関連する研究が立ち上がり分担研究者が主任研究者を務めることになった（厚労科研 成育疾患克服等次世代育成基盤研究事業 小児死亡事例に関する登録・検証システムの確立に向けた実現可能性の検証に関する研究）。研究を進めていくにつれ、研究成果が重複しうる状況となってきたため、本研究班の森研究班長と相談の上、最終年度は研究分担者から外れることとし、二か年の研究を終了とした。

医療データベースのリンケージ研究実施に際して適用される法令等に関する検討（分担：掛江直子）

本分担研究では、既存の医療情報を医学研究のために利活用するに際して、適用される現行法令や指針にはどのようなものがあり、また、各法令等にはいかなる事項等が規定されているのかにつき検討を行った。

個人情報の取扱いについて適用される法令は、各医療情報を保有する機関によって異なり、また、各法令の規定が一律ではない。利用目的、情報の内容、情報を保有する者や情報を取扱う主体によって適用法令や該当条文は異なり、また、情報の内容については、法の規制対象外の情報、法令の規制を受ける情報であっても、どの

ような質(種類)の情報であるか、すなわち、個人情報、要配慮個人情報、対応表により本人を識別することができる匿名化された個人情報、本人を識別することが不可能な匿名化された個人情報、統計情報であるか等によって、情報の取扱等にかかる適用法令や該当条文が異なっており、遵守事項に違いが生じる。また、法令が規定する個人情報の取扱の他に、日本の研究機関で実施され、または日本国内で実施される人を対象とする医学系研究については、医学系指針の遵守が必要となるが、当該指針の適用除外となる人を対象とする医学系研究もある。

個人情報たる医療情報の取扱いがこのような複雑な法制度となっていることにつき、運用・遵守しやすいシンプルなルールを作ることの必要性や、個人情報の取扱のあり方や考慮・保護すべき事項の再検討を含む抜本的な対応が必要である等の指摘がなされている。

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：森崎菜穂、Mahbub Latif)

本分担研究では、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、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。

リンケージ手法については、研究初年度に諸外国における人口動態統計間での連結手法について情報を収集し、それを参考に、2011年度に出生した児の出生票と死亡票をリンケージする手法を複数比較し、高精度の連結に必要な変数を選定した。研究2年目は、出生票と母の死亡票をオンライン登録情報を用いて連結する方法を提案し、研究3年目に本手法を2014-2016年度出生にも適応した。本手法により家族構成員の連結が可能となり、出産や中絶後の母の死亡のリスク因子の解明、更には兄弟の同定により同胞死亡についても背景因子の検討が可能となることが分かった。

一方で、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促進については、3年間を通して行い、約30本の英語原著論文を発表した。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。

さらに、研究3年目には康永分担班と共同することでDPC情報による周産期医療に関するレジストリ情報の代替可能性について検討した。既存の医師入力型の臨床レジストリに含まれている情報のうち

処置・投薬(病名以外)についてはDPC情報により代替できる可能性が充分高いこと、その際には施設名と患者番号が重要なリンケージ・キーになること、一般社団法人診断群分類研究支援機構を介在することでDPC情報を組織的に収集し臨床レジストリと連結するシステムが確立可能であることが示唆された。

D. 考察

本研究初年度は各分担班同士の情報共有を促し、各種データのリンケージ及び利活用の推進を行った。各分担班でも研究が進んでいるとともに、他分野のデータベースとの連結可能性についても模索が行われた。

研究2年目は、DPCデータベースや小児慢性特定疾病データベースなどの大規模データベースを他のデータとリンケージして活用する際に重要となる妥当性評価を行うとともに、平成15年より導入された人口動態統計オンライン報告システムに含まれている個人識別符号を利用して出生票とその母の死亡票を高精度にリンケージするなど、新たなリンケージ手法を検討した。

そして研究3年目は、研究1-2年目に活用してきた人口動態統計や周産期医療に関する臨床レジストリの継続的な利活用のみならず、レセプトデータベースやDPC情報へのリンケージの拡大を検討した。また、これらのデータとのリンケージ体制の構築と実務を通して、現状の可能性と問題点とまとめた。

3年間を通して、得られたデータベースの解析から、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、産科医・小児科医・疫学者とともに複数のエビデンスを数多く発表することができた。さらに、これらのデータとのリンケージ体制の構築と実務および医療情報の二次利用による臨床研究を行う際に留意すべき現行法令や指針とまとめることを通して、医療情報のデータリンケージとその利活用の可能性と問題点についてまとめることができた。

E. 結論

「医療分野の研究開発に資するための匿名加工医療情報に関する法律」(平成29年5月12日公布)の成立を受けて、日本でも医療データを他のデータベースと連結し活用することが促進されると思われる。

本研究では、データベースをリンケージし解析するという実践を通して、数々の臨床・社会的課題にそぐう解析を行い実際多くの研究論文の刊行に成功することで、データリンケージにどのようなその可能性が潜んでいるのかを示すことができた。

現在、他分野のデータベースとの連結可能性についても模索が行われている。今後、データ・リンケージにより質の高いエビデンスが産出できるようなシステム作りや、そのデータが活用できる分野の拡大が望ましいと思われる。