

196. データリンケージによる周産期疫学研究の推進

森崎 菜穂

*国立成育医療研究センター研究所 成育政策科学部

Key words : Probabilistic Data Linkage, 周産期疫学

緒 言

日本の周産期医療は大変進んでおり、妊婦死亡率および新生児死亡率は世界最低水準である。一方で、世界最低水準の出生率・不妊治療の増加、妊婦の高齢化およびやせ、経済的格差の拡大、など、特有の問題を多く抱えている。しかし、日本での周産期医療のデータの解析を適切な手段で行ってきた研究は極めて少なかった。このような状況において、実際の医療現場においては、日本の現状に即さないことがあったとしても、海外における研究結果が利用され、その結果各医師の経験則に基づいた医療が提供されてきた傾向があった。データに裏づけされた適切な根拠に基づく適切な医療を体系立てて患者に届けるためには、まず、日本に特有な問題を分析し治療法を選択するためのビッグデータを構築し、臨床に役立つように適切な解析を行う必要があった。

本研究を始めるまでの問題点としては、周産期医療に携わる産婦人科および新生児科の領分やの専門家が独立して研究を行うことが多かったこと、患者レジストリの運用はされていても適切な分析を行える人材不足のため十分にデータがエビデンス構築に寄与できていなかったこと、などが挙げられた。このため、本研究では、下記の4つの目標を持った。①日本産婦人科学会と日本未熟児新生児学会がそれぞれ固有に所有する周産期情報のデータベースをリンケージしたビッグデータを作成し今後の周産期疫学研究の基盤とする、これにより、いままで同じ母児の情報であるにも関わらず、産科が得意とする妊娠中の管理についての情報と、小児科が得意とする出生児の予後についての情報が連結され、包括的な周産期医療の評価が行えるようにする、②上記のデータベースの解析により日本の周産期医療の改善に直結する提言を多数算出し、妥当な根拠に基づく適切な医療に貢献する、③上記を通して、隣接する分野でありながら協力体制が希薄であった産婦人科医療関係者と小児科医療関係者の協力体制をはぐくむ、④周産期分野において、ビッグデータ解析に必要な疫学統計的知識や技術を持ちうる人材を育成する。

方 法

1. 利用したデータベース

日本産科婦人科学会周産期登録データベース (*Perinatal Research Network Database*: PRN database): 本データベースは1995年より継続的にわが国の全分娩の約16%について詳細な母児情報を収集しているわが国の唯一最大の周産期データベースである。施設参加は任意であるが、2011年には200を超える施設における全分娩について、母体情報、胎児情報および分娩周辺の関連医学情報を100項目以上、匿名化し、日本産科婦人科学会周産期委員会が保存蓄積している。今までに本データベースを用いて、子癇、妊娠糖尿病、多胎児の予後などについて研究がされてきた。

全国総合周産期センターネットワークデータベース (*Neonatal Research Network Database*: NRN database): 本データベースは日本未熟児新生児学会により、2003年より継続的にわが国のNICUにて加療を受けた極低出生体重児の約50%について詳細な母児情報および1.5歳、3歳の長期神経学的予後を収集しているわが国の唯一最大の極低出生体重児に関するデータベースである。参加施設は年々増加しており、2011年には165施設にわたる。今までに本データベースを用いて、超早産児の3歳児予後、出生前ステロイドの予後への影響、経静脈栄養の予後への影響、などの研究がされてきた。

我々は上記のそれぞれのデータベースの2004-2011年の全対象症例の情報を利用した。

2. リンケージの方法

NRN データベースおよび PRN データベースは院内カルテをもとに医療従事者により各施設で記入されており、これらの情報は個票レベルで連結が可能であると考えた。一方で、いずれのデータベースも手入力であるため入力ミスが一定頻度で見られること、またデータ項目が一部 PRN データベース (n=621,019)、NRN データベース (n=34,784) に共通する母児情報を用い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構成¹⁾を用いた。不完全データを含む情報の処理において Expectation-maximization (EM) アルゴリズムを利用することで maximum likelihood (ML) 計算によりモデルのパラメーターを計算した²⁾。この理論のもと、m-probabilities (1-エラー率) は各変数の信頼度を反映するように選定した。また μ -probabilities (ある値の出現頻度) は出生票における変数の分布から算出した。また、誤字や入力ミスが多いことが予想される出生体重に関しては、編集距離 (Levenshtein 距離) を用いて 2 値の一致度を評価した。また、マッチペアの信頼性を維持するために、マッチスコアのカットオフ値は規定値である 10 と設定した。

データのリンケージには LinkPlus Beta version (U.S. Centers for Disease Control and Prevention (CDC)、Cancer Division、Washington DC) を、その他の解析には Stata/MP version 12.0 (Stata Corp LP、College Station、Texas) を用いた。

3. 連結データベースの解析

第 50 回日本周産期新生児学会にてワークショップを開催し、連結されたデータベースおよびそれをもとにした解析について産科医および新生児科医に紹介し、同データベースを用いた研究提案を募集した。また、国立成育医療研究センターにて、若手の産科医および新生児科医を対象に、定期的に臨床研究セミナーを開催し、研究計画の立案、統計学的解析の指導、および論文執筆の指導を行った。

結 果

表 1 に 4 つのデータベースのマッチ率、そして母児情報のリンケージに利用した m-probability と、得られた連結データにおける情報の一致率を示す。

表 1. PRN データベースと NRN データベースのリンケージ

	PRNデータベース		NRNデータベース		PRNとNRNの
	m-probability	信頼性	m-probability	信頼性	データ一致率
性別	0.99	99.1%	0.99	99.2%	97.8%
生年	1.00	100.0%	0.99	99.8%	99.8%
誕生日	0.01	99.3%	---	-----	-----
母体年齢	0.90	95.3%	0.8	93.6%	90.9%
在胎週数	0.98	95.8%	0.99	99.6%	99.3%
在胎日数	0.99	98.7%	0.99	97.2%	97.4%
出生体重	0.95	98.3%	0.99	99.4%	96.7%
胎児数	---	-----	0.95	99.2%	-----
出生順位	0.95	99.9%	0.95	99.2%	98.4%
既往分娩数	---	95.7%	0.8	95.7%	94.5%

PRN データベースと NRN データベースの probabilistic record linkage 時に用いた m-probability、各データベース項目の信頼性 (出生票とのデータ一致率)、および両データベース連結時のデータ項目の一致率。

PRN データベースと NRN データベースのそれぞれにおいて、複数の出生票とマッチした症例、出生票とマッチしなかった症例、そしてマッチスコアが 10 未満の症例を除外したところ、PRN データベースの出生票との連結率は 99.1 % (615,941/621,019)、NRN データベースの出生票との連結率は 93.5 % (32,539/34,784) となった。また、連結された PRN データベースと NRN データベースのデータ項目の一致率は、96-99 % といずれも高かった。

考 察

今まで、固有識別番号を用いずに人口動態統計や臨床データベースを確率的リンケージ法により連結することにより研究に活用したケースは少なく、我が国においては初めての試みである。Hall ら³⁾ は米国オハイオ州南東およびケンタッキー州北部において、2007-2010 年に出産した初産婦に人口動態統計出生票・母の病院退院カルテ情報、そして産後の家庭訪問情報を確率的リンケージにより連結した。また、Herrchen ら⁴⁾ は米国カルフォルニア州において、1990-2000 年の人口動態統計出生票・死亡票と母の病院退院カルテ情報、児の生後 14 日以内の入院カルテ情報を確率的リンケージにより連結し、このデータを活用し Schmitt⁵⁾ は周産期医療にまつわる医療費の大部分が早産児や低出生体重児など一部の児によるものであり早産の予防が州の医療費削減に約立つことを報告し、そして Burgos⁶⁾ は入院加療を要する新生児黄疸の年次推移を報告しさらにリスク因子の解析を行った。

我々が作成した連結データベースを用いて、平成 28 年 5 月現在、妊婦の体格と分娩時合併症のリスクの関連、母体の社会的背景と児の予後、胎内一児死亡双胎妊娠における分娩時期の検討、日本人にとって最適な妊娠中の体重増加量の検討、などの解析を若手の産科医・小児科医・社会疫学者が連携して行い、学会発表を 10 本以上行っており、また学会発表に続き論文化もなされている。今まであまり見られなかった分野を超えた協力のもと、今後も我が国の周産期医学の向上に貢献する多くの研究成果が期待される。

共同研究者

本研究の共同研究者は、国立成育医療研究センターの森臨太郎、左勝則である。また、学会データベースの利用許可および指導について日本成育新生児医学会（旧日本未熟児新生児学会）会長楠田聡、日本産婦人科学会周産期委員会会長斎藤滋に、また疫学研究セミナーに参加し、現在当データベースの解析に関わっている産婦人科および小児科の先生方に、そして本研究を支えて下さった上原記念生命科学財団の関係者に謝辞を申し上げたい。

文 献

- 1) Fellegi IP, Sunter AB. A Theory for Record Linkage. *Journal of the American Statistical Association*. 1969;64(328):1183-210
- 2) Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977;39(1):1-38
- 3) Hall ES, Goyal NK, Ammerman RT, Miller MM, Jones DE, Short JA, Van Ginkel JB. Development of a linked perinatal data resource from state administrative and community-based program data. *Matern Child Health J*. 2014 Jan;18(1):316-25. doi: 10.1007/s10995-013-1236-7. PMID: 23420307.
- 4) Herrchen B, Gould JB, Nesbitt TS. Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Comput Biomed Res*. 1997 Aug;30(4):290-305. PMID: 9339323.
- 5) Schmitt SK, Sneed L, Phibbs CS. Costs of newborn care in California: a population-based study. *Pediatrics*. 2006 Jan;117(1):154-60. PMID: 16396873.
- 6) Burgos AE, Schmitt SK, Stevenson DK, Phibbs CS. Readmission for neonatal jaundice in California, 1991-2000: trends and implications. *Pediatrics*. 2008 Apr;121(4):e864-9. doi: 10.1542/peds.2007-1214. PMID: 18381515.