

令和 6 年 5 月 24 日現在

機関番号：34504

研究種目：若手研究

研究期間：2021～2023

課題番号：21K13274

研究課題名（和文）死因の聞き取り調査における質問項目の重要度の推定

研究課題名（英文）Estimation of the importance of survey items in verbal autopsy for cause of death

研究代表者

國濱 剛（Kunihama, Tsuyoshi）

関西学院大学・経済学部・准教授

研究者番号：40779716

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究課題では、死因分析における聞き取り調査表の質問項目の重要性を定量化する統計手法を考案し、実際にデータから各々の重要度を推定するための計算アルゴリズムを開発した。加えて、提案手法を実際の死因聞き取り調査データを用いた実証分析へと応用し、症状・病歴などの質問項目の重要度の推定を行った。分析結果として、早産や低体重などの要因に加え、人口学的情報である年齢・性別が死因との関連性が高いことを明らかにした。

研究成果の学術的意義や社会的意義

先行研究において個々の質問項目と死因の関連性の強さについて議論が行われてきたが、本研究課題では、他の質問との情報の重複を除外した上で、個々の質問項目が死因に関して持つ固有の情報量を計算した点に独自性がある。提案手法を活用することで、今後行われる死因の聞き取り調査において、質問表から重要度の低い質問項目を減らす、逆に重要度の高い症状に関しては新たな項目を加えることにより、死因分布の分析精度が向上する可能性がある。

研究成果の概要（英文）：This research project developed a statistical method to quantify the importance of survey items in the analysis of distributions of cause-of-death and created a computational algorithm to estimate the importance of each item from the data. Additionally, we applied the proposed method to empirical analysis using actual cause-of-death verbal autopsy survey data and estimated the importance of survey items such as symptoms and medical history. The analysis revealed that, in addition to factors like prematurity and low birth weight, demographic information such as age and sex has a relatively strong association with the cause of death.

研究分野：ベイズ統計学

キーワード：ベイズ統計学

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

死因は公衆衛生政策の根幹を成す人口学的情報であるが、一部の発展途上国では人口動態統計の整備が不十分であることに加え、医療機関外での死亡数が多いこともあり、地域全体を網羅する死因情報に大きな不確実性が残る。そこで、遺族に対して聞き取り調査を行い、故人の症状や病歴などから死因を特定する口頭剖検が用いられている。口頭剖検のための調査票には多数の質問項目が含まれるが、すべての項目が死因分析において等しく重要であるとは限らない。聞き取り調査を効率的に実施するため、死因との繋がりが弱い項目や余分なものを調査票から外す、逆に死因との関連性が高い症状に対しては質問項目を増やす等が考えられるが、そのためには何を基準に質問項目の重要性を測るのか、またその重要度をどのように実際のデータから統計的に推定するのかについて明らかにする必要がある。

2. 研究の目的

本研究課題の目的は、死因特定のために実施される遺族への聞き取り調査において、調査票に含まれる故人の症状・病歴・生活習慣などの質問項目がそれぞれどの程度重要であるのかを明らかにすることである。死因の聞き取り調査には質問項目数が多いことに加え、遺族が症状の一部を覚えていないため個票データに多く欠損が存在するなどの特徴があるが、本研究ではそれらを考慮に入れたベイズ統計手法を新たに考案する。そして、提案手法を実際の聞き取り調査データを用いた実証分析に応用し、各々の質問項目の重要度を統計的に推定する。

3. 研究の方法

ベイズ統計手法を用いて死因と質問項目の同時確率分布のモデル化を行い、その同時分布に基づいて死因に対する質問項目の重要性を数値化することを考案した。具体的に、情報理論の概念として広く応用されている相互情報量と条件付き相互情報量の二つを重要度の指標として採用する。相互情報量は個々の質問項目から得られる死因の情報量を表し、この値が大きいほど重要度が高いと解釈できるが、他の質問項目が持つ情報との重複を考慮に入れていないため、該当項目の重要性を過大評価する恐れがある。一方で、条件付き相互情報量は、個々の質問項目が固有の死因情報をどれだけ持っているかを表しており、この2つの指標を参考にすることで質問項目の重要度をより正確に表現することができる。提案手法に対して、マルコフ連鎖モンテカルロ法に基づく計算アルゴリズムを開発し、聞き取り調査データを使って各々の質問項目の重要度を推定するベイズ統計手法を考案した。最終的に、実際の死因聞き取り調査の Population Health Metrics Research Consortium (PHMRC) データを使った実証分析を行い、具体的にどの質問項目の重要度が高いのかを明らかにする。

4. 研究成果

提案した統計手法を実際の死因聞き取り調査データによる実証分析に応用した。具体的に、1619人の新生児に対する5個の死因カテゴリー（出生時仮死、先天的形成不全、髄膜炎・敗血症、肺炎、早期産）と97個の質問項目から構成される PHMRC 新生児データを分析に用いた。考案した重要度の指標の計算には死因と質問項目との同時確率分布が必要となるが、ベイズ因子モデルを採用して最適な因子数はクロスバリデーションを通して選択を行った。分析の結果、本研究の主な成果として図1と図2が得られた。

図1は基準化相互情報量の事後平均値の大きな質問項目リストを表している。最初の二つは体重と身体的大きさで、死因が持つ情報量の約10%を占めている。また、早産、晩産、身体的異常、啼泣・授乳に関連する症状が高い値を示している。人口学的情報に関して、年齢はリストにあるが性別は入っていない。図2は基準化条件付き相互情報量の事後平均値の大きな質問項目リストを表している。まず、図2の条件付き相互情報量の推定値は図1の基準化相互情報量の推定値より全般的に小さくなっているが、これは個々の質問項目が持つ死因の関連情報に重複が存在することを意味する。体重、身体的大きさ、早産、晩産、身体的異常、啼泣・授乳に関連する症状など、図1で比較的大きな相互情報量を持つ質問項目の多くが、条件付き相互情報量の指標においてもリスト内に入っていることがわかる。逆に、図1と図2を比較すると違いもあり、一番大きいものが性別の位置であった。図1には表示されていないが性別の相互情報量は非常に小さい一方で、図2にあるように条件付き相互情報量の事後平均値では他の質問項目より大きな値を示している。条件付き相互情報量が相互情報量より大きくなっているのはすべての質問項目の中で性別だけであった。つまり、新生児の性別のみが単独で与えられても死因に関する情報は直接的にはあまり得られないものの、他の質問項目を通して間接的には固有な死因情報をもたらすことを意味している。

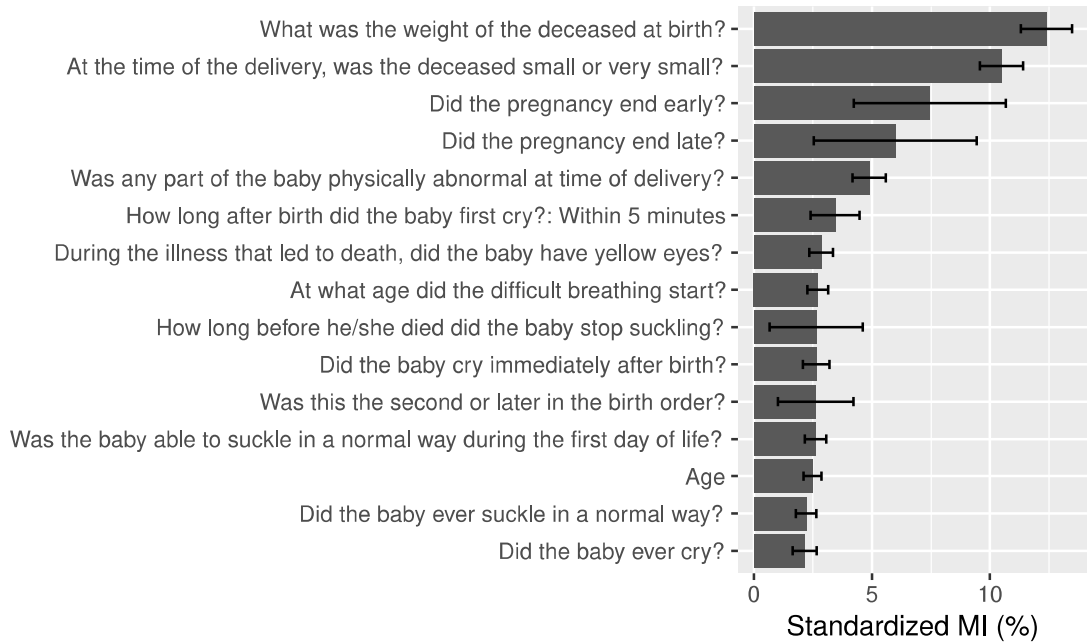


図 1 : 基準化相互情報量の事後平均値の大きな質問項目リスト。バーは標準偏差に対応する。

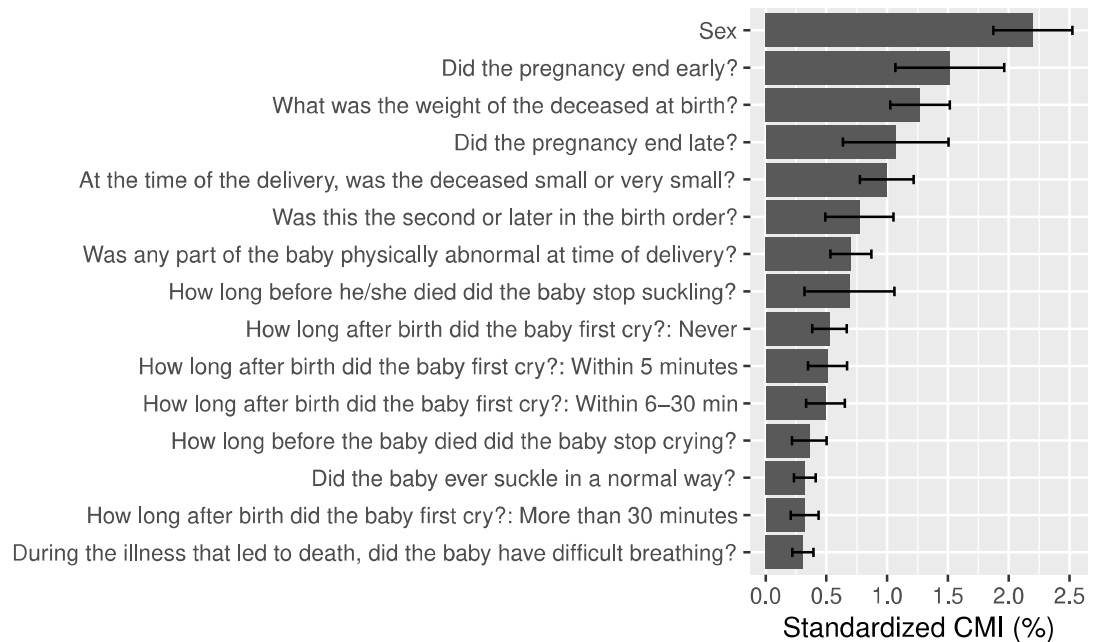


図 2 : 基準化条件付き相互情報量の事後平均値の大きな質問項目リスト。バーは標準偏差に対応する。

関連分野における先行研究の多くは、死因の予測精度を向上させるために質問項目が与えられた下での死因の条件付き分布の開発に主な焦点を当てており、説明変数の重要度についてそれほど多くは議論されていない。一部の研究論文において、個々の質問項目と死因の関連性の強さについて個別に分析を行っているが、他の質問項目との情報の重複については考慮に入っていない。本研究課題では、他の質問との情報の重複を除外した上で、個々の質問項目が死因に関して持つ固有の情報量を計算した点に独自性があり、国内外問わずに関連分野に貢献があると考えている。例えば、今後実施される死因の聞き取り調査に対して、質問表から重要度の低いような質問項目を減らす、逆に重要度の高い症状に関しては新たな関連項目を加えることにより、聞き取り調査の時間を短縮しながらも死因分布の推定精度を向上できる可能性がある。

今後の展望としては、複数の方向性が考えられる。まず、他の死因聞き取り調査データを用いた実証分析に提案手法を用いて比較を行うことである。例えば、PHMRC 成人データにおいても性別は本研究結果と同様に、他の質問項目とは異なる死因との関係性を持つのか等を調べることが考えられる。また、PHMRC 成人データは死因カテゴリー数も質問項目数も、本研究課題で用いた新生児データのものよりも多いため、推定時間の短縮のためにマルコフ連鎖モンテカルロ法を用いたパラメータ推定方法の更なる効率化が必要になるかもしれない。加えて、得られた重要度を用いた変数選択を行う統計手法の開発も考えられる。本研究課題で考案した重要度は基準化することで 0 から 1 までの値を取るが、実際にどの程度小さい値だと分析に必要ないと判断できるのか現時点では明確ではない。今後の課題として、重要度の低い順に質問項目を分析リストから外すことで死因分布の推定精度がどのように変化するかなどを詳しく分析する必要がある。さらに、提案手法では死因と質問項目との同時確率分布を必要とするが、この同時確率分布自体の精度を向上させることも考えられる。実際のデータにおける死因と質問項目との真の関係を柔軟に捉えることができる同時確率分布を構築できれば、それに基づいて質問項目の重要度を測ることで、より正確な推定値を得ることができると考えられる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 國濱剛
2. 発表標題 Bayesian analysis of verbal autopsy survey using multivariate probit model with age and sex-dependent association between symptoms
3. 学会等名 統計学の理論と応用のフロンティア
4. 発表年 2022年

1. 発表者名 Tsuyoshi Kuniyama, Zehang Richard Li, Samuel J. Clark, Tyler H. McCormick
2. 発表標題 Bayesian analysis of verbal autopsy data using probit model with age- and sex-dependent association between symptoms
3. 学会等名 The 6th International Conference on Econometrics and Statistics (EcoSta 2023) (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
米国	University of California, Santa Cruz	Ohio State University	University of Washington