

科学研究費助成事業 研究成果報告書

令和 2 年 6 月 9 日現在

機関番号：82801

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K09222

研究課題名(和文) ベイジアンネットワークを用いた知識学習型接触者健診モデルの構築

研究課題名(英文) Construction of Knowledge Learning Model for Tuberculosis Contact Investigation Using Bayesian Network

研究代表者

内村 和広 (Uchimura, Kazuhiro)

公益財団法人結核予防会 結核研究所・臨床・疫学部 疫学情報室・副部長

研究者番号：30247283

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：健診対象者の属性情報をもとにした接触者健診モデルをベイジアンネットワークの手法を用いて構築し、接触者健診結果情報からの学習により、IGRA陽性者の感染確率についての推定を事前確率のかたちで推定を行なった。属性情報について情報なしを基準とした場合と情報を学習した場合のIGRA陽性患者の真の感染確率を推定し比較した。学習情報として情報価値の高いものは初発患者有空洞情報、接触タイプ、接触濃厚度、接触者では外国出生の項目となった。用いたシミュレーションデータは集団感染事例が含まれており、接触者対象集団の感染率は比較的高いと考えられるが、推定された感染確率としての事前確率では事前確率の上昇がみられた。

研究成果の学術的意義や社会的意義

一部実データには情報なしの項目等があるため、現時点では補完したシミュレーションデータによる解析でモデル精度向上をすすめているが、実情報量が充足するにつれ、より現実的分析に近づくと思われる。また、諸項目の入力後(あり/なし)のIGRA陽性確率(事後分布として)の推定も接触者健診実施上の有効情報になると思われる。

研究成果の概要(英文)：A Tuberculosis contact investigation model based on the characteristics of the index tuberculosis case and contacts was constructed using the Bayesian network method. And the probability of infection of IGRA-positive people was estimated by learning from contact investigation result. Regarding the attribute information, the infection probability (true infection probability) of the IGRA-positive patients when the information was used as a reference and when the information was learned was estimated and compared. The items with high information value as learning were items of initial patient cavity information, contact type, contact density, and contact items for foreign births. The simulation data used include outbreak cases, and the infection rate of the contact target population is considered to be relatively high. However, the estimated a priori probability of infection is about 1.2 to 1.7 times the prior probability.

研究分野：結核疫学

キーワード：結核 接触者健診 ベイジアンネットワーク

1. 研究開始当初の背景

日本において、結核は1年間に1万8千人以上の患者が新たに発生し、2千人に近い死亡を引き起す最大の慢性感染症であり、けっして過去の病気とはなっていない。結核対策で最も重要なもののひとつが接触者健診であり、その重要度は増している。世界保健機関(WHO)は2035年までに世界規模での結核低蔓延化を目標とする結核終息戦略(End TB Strategy)を発表し、3つの大きな戦略上の柱のうちの「統合化された患者中心の予防と治療」において、接触者健診をその第1に位置付け、拡大戦略を図っている。

一方、近年の人工知能の発展とともに、ベイジアンネットワークと呼ばれる手法の応用が広がっている。ベイジアンネットワークとは、統計的にはベイジアン理論を持った、不確実性を含む事象の予測や合理的な意思決定、観測結果から原因を探る障害診断などに利用することのできる確率モデルの手法である。ベイジアンネットワークの大きな利点に、知識の学習機能により推定結果の精度を向上していく点があげられる。

このベイジアンネットワークの手法を、接触者健診の評価と健診拡大の有無の意思決定に応用することが考えられる。接触者健診実施数とそこから発見された患者数や、IGRAによる結核菌感染陽性者数などを、初発患者、および接触者の属性別による条件付き確率の情報として蓄積することで、新たに行なわれる接触者健診での拡大の意思決定を、健診を拡大した場合に発見される患者や結核菌感染陽性者数を様々な条件下での事後確率として推定される。先に触れたように、個々の保健所での集団感染対応事例は散発的であるが、これを多数の保健所での事例を集積する、すなわち接触者健診ベイジアンネットワークモデルに学習させることにより、新たな接触者健診対応が必要になった保健所で、過去に対応がなかったような初発患者・接触者集団においても、健診拡大時の結果の推定を行なうことが可能となる。

2. 研究の目的

結核低蔓延化を目指す日本はもちろん、世界的な結核対策戦略においても結核患者の接触者健診の効率的実施と健診結果の評価の意義は大きい。しかし、実際の健診実施においては健診対象者の拡大の意思決定や、個々の保健所では集団感染事例数が少なく結果の評価の精度が低いことが課題としてあげられる。本研究では、知識学習型理論をもとにしたベイジアンネットワーク理論を用いて接触者健診モデルを構築し、複数保健所での集団感染事例の情報を構築したネットワークに学習させることにより、これらの課題を解決し、今後の日本および同様な結核疫学状況下の国々での(1)接触者健診の効率的な実施と(2)健診結果に対する根拠ある評価を可能とすることを目的とする。

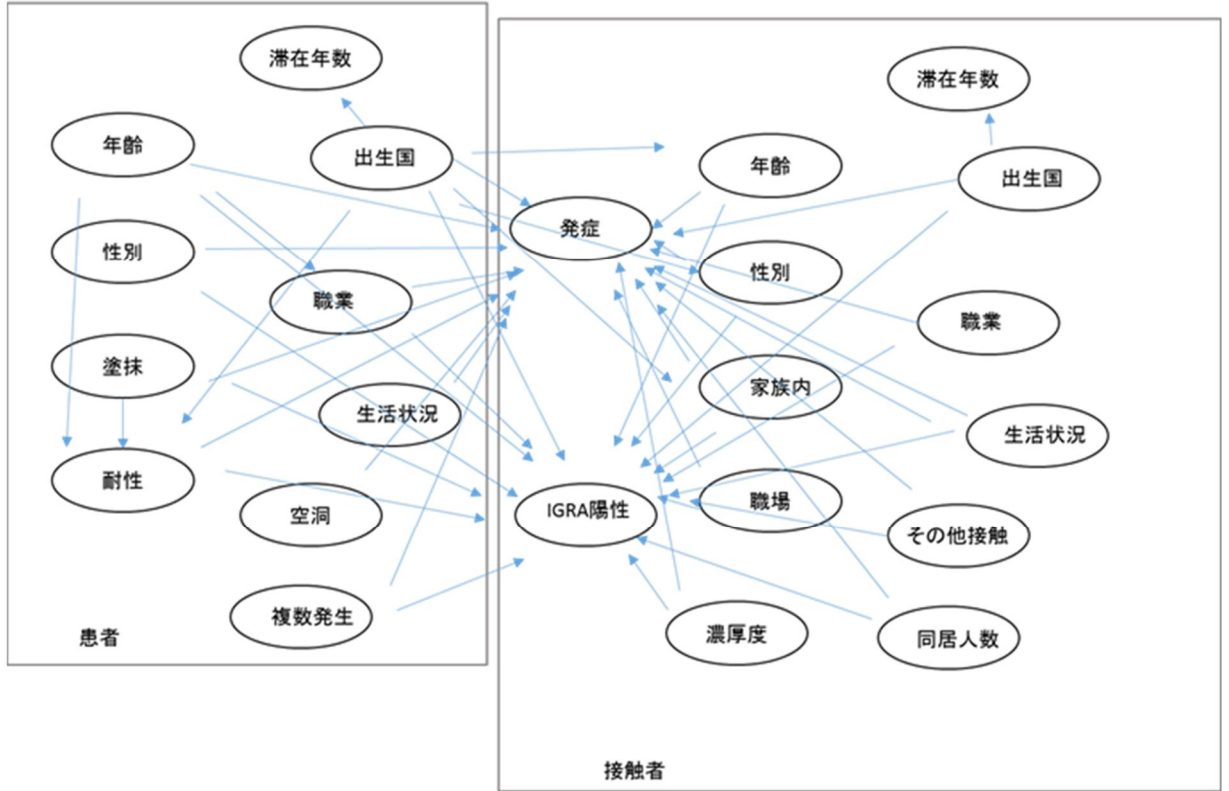
3. 研究の方法

健診対象者の属性情報をもとにした接触者健診モデルをベイジアンネットワークの手法を用いて構築し、保健所における接触者健診結果情報からの学習により、IGRA陽性者の感染感染確率についての推定を事前確率のかたちで推定を行なった。ベイジアンネットワークの初発患者ノードには、性、年齢階級、出生国(外国出生)、塗抹検査結果、胸部レントゲン有空洞、受診の遅れ、喫煙を設定した。接触者のノードには、IGRA結果、性、年齢階級、接触タイプ(家族、その他)、喫煙、接触濃度(現在のところ接触時間や接触空間等から便宜的に濃厚、非濃厚)を設定した。分析はBAYONET(NTT数理データシステム)上で行なった。

4. 研究成果

接触者健診のベイジアンネットワークを初発患者およびその接触者の各パートにおいて図 1 のように構成した。

図 1：接触者健診のベイジアンモデル構成図



また、学習効果の推定のため、保健所データの一部未収集項目データを仮想的に補完したシミュレーション用データを用意した。

そこでのパラメータ設定を以下の表 1,2 にあげる。

表 1 初発患者設定パラメータ

シミュレーションデータの生成 - 主なパラメータ

接触者

高蔓延国出身（濃厚度を同出身か否かで 1 : 10）
それ以外

性（濃厚接触を初発 男 接触者 男 1）
濃厚接触を初発 女 接触者 女 1
濃厚接触を初発 男 接触者 女 0.5)

年齢（30歳以上で既感染率を1.5倍
濃厚接触度を同年齢層か否かで1 : 1.5)

職業（学生、常勤、無職
濃厚接触を同職業か否かで1 : 1.5)

表2 接触者設定パラメータ

シミュレーションデータの生成 - 主なパラメータ

接触者

初発の塗抹 (陽性か陰性 感染 : 非感染 5 : 1)

初発の空洞 (ありかなし 感染 : 非感染 5 : 1)

生活状況 (困難か否か 感染 : 非感染 2 : 1)

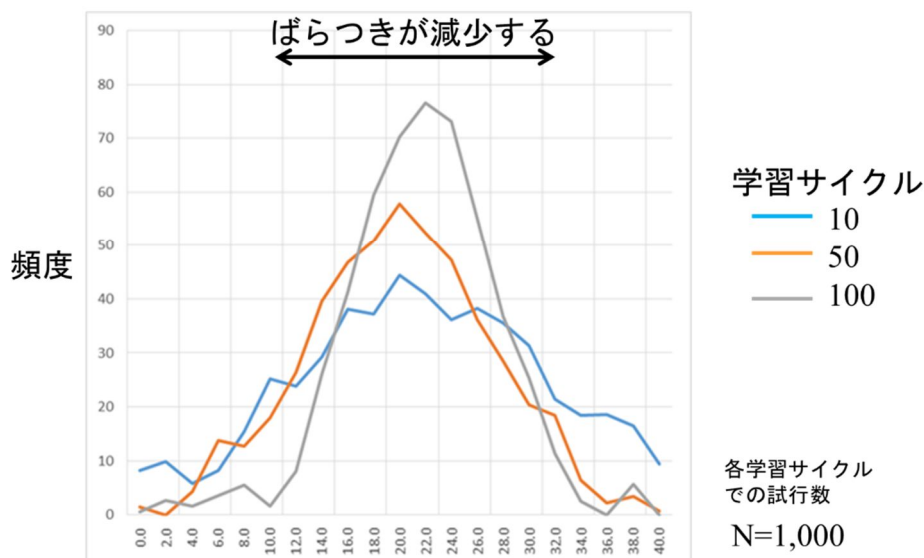
濃厚接触 (感染 : 非感染 5 : 1)

感染者の各属性をパラメータに誤差を与えて(正規分布)を生成して、高蔓延国出身接触者 10 人を平均とする乱数を発生して接触者側のシミュレーションデータとした。学習過程を、接触者が高蔓延国出身か否かを含む初発患者側のネットワーク各ノード状態に対する IGRA 陽性率のデータを、高蔓延国出身接触者の感染率 20%で発生させた 20 人の接触者健診シミュレーションデータとして、これをくりかえし与えた。各パラメータの事前分布の初期値は無情報事前分布とし、各サイクルでパラメータの事前分布を推定し、これを学習とした。さらにサイクル数を 20, 50, 100 で推定を行なった。

属性情報について情報なしを基準とした場合と情報を学習した場合の IGRA 陽性患者の感染確率(真の感染確率)を推定し、比較した。学習情報として情報価値の高いものは初発患者有空洞情報、接触タイプ、接触濃厚度、接触者では外国出生の項目となった。用いたシミュレーションデータは集団感染事例が含まれており、接触者対象集団の感染率は比較的高いと考えられるが、推定された感染確率としての事前確率では約 1.2~1.7 倍程度の事前確率の上昇がみられた。初発患者有空洞情報、接触タイプ、接触濃厚度を無情報、外国出生のみありとした場合と初発患者有空洞情報、接触タイプ、接触濃厚度ありとした場合は $0.89 / 0.52 = 1.7$ であった。接触濃厚度、外国出生を無情報、初発患者有空洞情報、接触タイプをありとした場合から、接触濃厚度、外国出生をありとした場合は $0.89 / 0.61 = 1.4$ であった。(図2)

図2

結核高蔓延国出身接触者のIGRA陽性者の事後分布から推定した感染率



考察

学習過程による事前情報の蓄積により、接触者健診の出力値の事後分布の精度向上が期待できる。接触者健診実データによる検証を行ない、今後もひきつづき、ベイジアンネットワークモデルの改良を進める必要がある。保健所での実地応用が極めて重要であるが、新型コロナウイルス感染症の流行で保健所データ応用のまとめが遅れており再開後 2020 年内に論文として早急にまとめる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔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
研究分担者	河津 里沙 (Kawatsu Lisa) (10747570)	公益財団法人結核予防会 結核研究所・臨床・疫学部・研究員 (82801)	