

厚生労働科学研究費補助金（がん対策推進総合研究事業）

分担研究報告書

都道府県がん登録の全国集計データと診療情報等との併用・突合によるがん統計整備
及び活用促進の研究

研究分担者 加茂 憲一 札幌医科大学 医療人育成センター 准教授

研究要旨

研究テーマ「統計解析による都道府県がん登録データの活用」において、全国と地域（本報告では北海道）の罹患数の登録状況の比較を試みた。具体的には、北海道に着目し、MCIJとの比較検討を行った。登録率の比較においては、MI比-DCN割合の関係性を用いた回帰モデルによる登録の完全性補正を行った。2014年のMCIJと北海道がん登録報告におけるMI比とDCN割合を用いて、登録漏れを含む「真の罹患数」を推定した。推定結果と実測値の特性に着目し、両者の比較検討を行った。登録率に関しては、全国で約90%、北海道で焼く80%と10%の開きがあった。数理的には、公差検証法によるモデル選択を行っており、全国では線形モデル、北海道では非線形モデルが選択された点に関する差異があった。推定された登録率に着目すると、全国については登録の完全性はかなり改善されており、そのことが登録の完全性に関する変数の一極集中を招き、その結果として登録の完全性以外の部分（地域特性・・・部位の分布やリスク要因）が際立つ結果であった。一方で北海道に関しては、二次医療圏ごとの登録状態に依然大きな開きが観察された。更には全国に比して地域特性にある程度の一様性が期待できることもあり、完全性を変数とした回帰モデルに適した状況であることが観察された。ただ二次医療圏の規模に関して突出する地域が存在し、それが推定に悪影響を与えている可能性は否定できず、数理的な工夫が必要であると考えられる。

A. 研究目的

全国がん登録の開始に関連して、がん登録に関する様々な環境が整備されてきた。特に登録の完全性に関しては、推奨される基準を充足する地域が年々増えてきている。しかしながら、未だにDCNやDCOが存在することから、罹患の登録漏れはゼロではなく、その分が過小評価されている危険性がある。この問題点を数理的に解決するために、登録の完全性に関する指標 (MI比とDCN割合) に基づいた回帰モデルを用いて登録率を推計する試みがなされてきた。

一般的に、回帰モデルは変数が一極集中していない状態が理想的である。しかし上記の通り登録の完全性が改善されてきたことが、完全性の指標の一極集中化に繋がり、回帰モデルを適用し辛い状況を引き起こしている。本モデルにおいては、完全性に関する挙動を回帰式で表現し、回帰式で説明できない部分を誤差と考えている。この考え方だと、説明変数となる完全性指標 (DCN割合) が低い値でかたまってしまう一方で、その他の誤差部分は依然として残ってしまい、結果として誤差部分の変動が変数の変動に比して大きくなる傾向となる。誤差部分には地域固有の変動部分が内包されているとも感ぜられる。もちろんDCN割合が低い値にかたまるのは望ましい状況ではあるが、適用する数理モデルに一考を要する状況である。同一地域 (本研究では北海道) においては固有の変動部分が小さい (地理的な一様性を有する) と考えられる点に着目し、全国と比較した結果を報告する。

B. 研究方法

地域がん登録における登録の完全性を補正する方法として、完全性の指標であるMI比とDCN割合との関連性を回帰構造により評価する方法が用いられてきた。具体的には、MI比を「罹患から死亡が発生する確率」と捉えることにより、罹患数が二項分布に従うと仮定することができ、死亡確率に対して回帰モデルを導入することができる。具体的には、MI比はDCN割合の影響を受けて変化する (高DCN割合においてMI比は1に近づく) 仕組みとなっているため、DCN割合を説明変数とするロジスティック回帰モデルが構築できる。ここでMI比とDCN割合の関連性については、線形構造と非線形構造の2種類が提案されているため (非線形が線形を含む数式になっている)、それらの優劣は公差検証法に基づいて決定する。回帰モデルに含まれるパラメータのうち「DCN割合が0%でのMI比」の部分が推定されれば、それに死亡数 (死亡数は全数把握されていることを用いる) を乗ずることにより、DCN割合がゼロ、すなわち登録が完全な状態であると見做される罹患数を推定することが可能となる。更に罹患数報告値との比をとることにより、登録率が推定される (参考: Kamo et al. (2007) Jpn J Clin Oncol)。

(倫理面への配慮)

特に必要なし

C. 研究結果

「B. 方法」に記述した手法を、2014年データに適用した。一つはMCIJ全国推計に用いられた41地域（基準を充足する地域）からの全国値再推計を行った。もう一つは北海道の二次医療圏（21の医療圏）のデータを用いて全道の罹患数を再推定した。図1に全国における推定結果、図2に全道における推定結果を示す。どちらのグラフも横軸DCN割合、縦軸MI比に対して、地域プロットを円で表している。比較のために図1と図2の縦軸と横軸は統一している。円の面積は罹患数の規模に比例している。一方でプロット上には線形モデルと非線形モデルの推定結果を、線形は直線で、非線形は破線で記している。

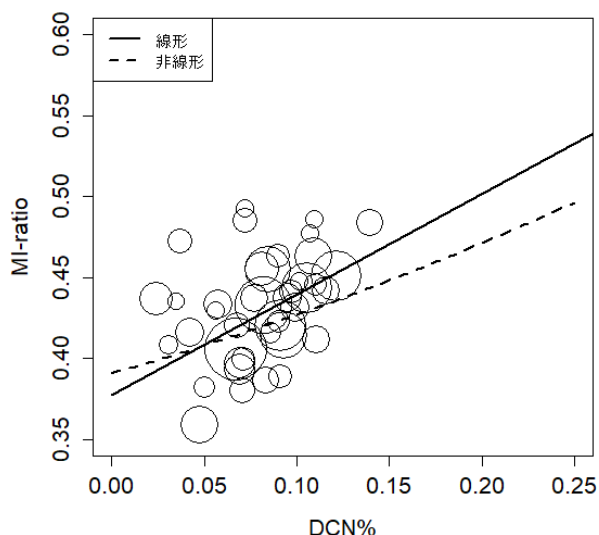


図1. 41地域から全国を推計

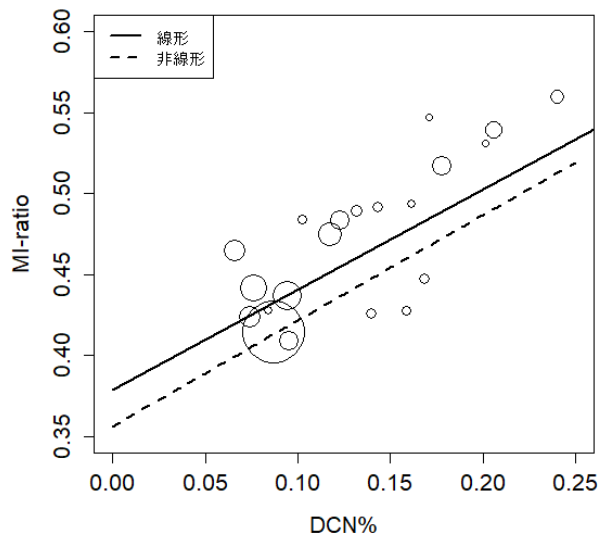


図2. 21の二次医療圏から全道を推定

公差検証法により、全国については線形モデル、全道については非線形モデルが選択された。それぞれのモデルを元に推定された罹患数は、全国で975053人、全道で53186人であり、報告値と比較すると登録率は全国で89.9%、全道で79.5%と推定された。北海道の登録率は全国平均に比べて約10%低い可能性が示唆された。

D. 考察

2014年におけるMCIJ報告の全国罹患数は、真の罹患数に対して約90%が登録されている結果が推定された。これまで推定されてきた登録率は80~85%であったことを鑑みると、登録の完全性はかなり改善されていることが伺える。一方で北海道に関しては、約20%の登録漏れの可能性が推定され、全国に比して

10%程度低い完全性であることが示唆された。ただし、推定で用いるモデルを公差検証法で選択したところ、全国で線形、北海道で非線形と異なるモデルが選択されたことに注意が必要である。これまでの解析（全国）においては概ね線形モデルが選択されてきたが、今回初めて北海道の解析に用いると非線形が選択されたのは興味深い。もし全国と同じ線形モデルを用いた場合、北海道の登録率は85.3%と推定された。モデル依存の過小評価分はあるにしても、全国に比しての登録率は低いものであった。北海道において非線形モデルが選択された原因としては、後にも述べる二次医療圏の規模において札幌圏が突出しているため、一つの二次医療圏のデータに全体の推定が引っ張られ、結果としてモデルとしては「柔らかい」非線形が選択された可能性が考えられる。

E. 結論

回帰モデルを用いて登録の完全性を補正する方法について、登録の完全性が高い全国と、北海道に適用し、結果に関する比較を行った。全国に関しては登録率が約90%と非常に高く、実測値のプロットからも、罹患数は完全性の影響よりも、それ以外の地域変動（部位の分布や地域依存のリスク要因）の方が強いことが目視できた。従って、これまでのように登録の完全性単独で回帰モデルを適用することには限界に近づいている印象を受けた。今後は地域差異の部分も包括的に評価できる統計モデルの開発が期待される一方で、依然としてDCNが存在することから完全性に関して

も同時に補正する手法の開発が必要であると考えられる。

一方で、21の二次医療圏から北海道全体の罹患数を推定するという状況に関しては、依然として完全性に医療圏間の差異が大きいこと、そして同じ北海道内ということで地域独自のばらつき要素が小さいことが実測のプロットから観察される。これはまさに登録の完全性に関する回帰モデルが適用しやすい状況であり、推定結果も妥当なものであった。回帰曲線は、実測のMI比に対して低めの推定結果が得られているようにも見受けられる。これは、北海道における札幌圏の人口規模が突出していることに推定が引っ張られていることが考えられる。今回の数理モデルは二項分布がベースとなっているため、地域の規模も反映されているが、それでも札幌圏の突出具合が非常に高い。今回の推定結果から、この突出が全体の推定に過度な影響を与えている危険性も考えられ、推定方法自体に一考の余地があると考えられる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

1. Matsuda, K.Saika, R.Tanaka, Y.Ito, K.Fukui, K.Kamo. Simulation models in gastric cancer screening: a systematic review. Asian Pacific Journal of Cancer Prevention, 19 (12), 3321-3334, 2018.

2. K.Iesato, T.Hori, Y.Yoto, M.Yamamoto, N.Inazawa, K.Kamo, H.Ikeda, S.Iyama, N.Hatakeyama, A.Iguchi, J.Sugita, R.Kobayashi, N.Suzuki, H.Tsutsumi. Long-term prognosis of patients with HHV-6 reactivation following allogeneic HSCT, *Pediatrics International*, 60, 547-552, 2018.
3. K.Tanaka, T.Kajimoto, T.Hayashi, O.Asanuma, M.Hori, K.Kamo, I.Sumida, Y.Takahashi, K.Tateoka, G.Bengua, K.Sakata, S.Endo. An in vitro verification of strength estimation for moving an ¹²⁵iodine source during implantation in brachytherapy, *Journal of Radiation Research*, 59 (4), 484-489, 2018.
4. K.Kamo. A new approach to classify growth patterns based on growth function selection and k-means method, *FORMATH*, 18(DOI: 10.15684/formath.18.003), 2019.
5. R.Tanabe, K.Kamo, K.Fukui, S.Imori. Statistical inference for estimating the incidence of cancer at the prefectural level in Japan, *Jpn J Clin Oncol* (accepted).

2. 学会発表

加茂憲一. 全国がん罹患数推定値における登録率の推定, 日本がん登録協議会第27回学術集会 : 2018年6月13-15日 : 沖縄.

H. 知的財産権の出願・登録状況

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし