

混合効果モデルによる都道府県別がん罹患数の区間推定

大阪大学大学院基礎工学研究科 田辺 竜ノ介
札幌医科大学医療人育成センター 加茂憲一
大阪大学大学院基礎工学研究科 伊森晋平
大阪国際がんセンターがん対策センター 福井敬祐

1. 導入

がん登録はがんの診断、治療、経過などに関する情報を集め、がん対策の立案・評価に寄与することを目的として行われる。がん登録による罹患数調査においては、登録漏れ等により報告された罹患数と実際のがんの罹患数が異なる可能性がある。従って、がん登録に基づく罹患状況の把握では、真の罹患数のうちどの程度が把握出来ているかという完全性が重要な要素である。完全性の算出には真の罹患数情報が必要であり、例えば Kamo et al.(2007) では回帰モデルによる完全性の補正を行い、真の罹患数を推定している。先行研究では全国値の完全性を推定する方法を提案しているが、これらを都道府県個別での議論に適用することはできない。よって本報告では、都道府県ごとに報告される罹患数の信頼度を計るために、罹患数の信頼区間を構成する方法を提案する。

本研究では、混合効果モデルを用いて罹患数を区間で推定する方法を2種類提案する。1つ目は、完全性の補正を行った上で各地域を独立にベイズ統計を用いて信頼区間の構成を行うもの。2つ目は、各地域の完全性の補正を行った上で、全国推計の手法における都道府県共通のパラメータ設定を使用し混合効果モデル及びベイズ統計を用いて各地域の罹患数の信頼区間を求めるものである。

2. 手法 がん登録においては罹患数は完全に把握されていないが、がんによる死亡数は完全把握されている。このとき、地域番号を i として、がん死亡数 m_i からがんの罹患数 n_i を推定する問題は二項分布 $\text{Bin}(n_i, p_i)$ におけるパラメータ n_i の推定問題として扱える。本研究の目的はベイズ統計を用いて n_i の信頼区間の構成と言い換えることが出来る。客観事前分布として n の事前分布に $\pi(n_i) \propto 1/n_i$ を置く。また完全性の補正として本研究では死亡確率を $p_i = \alpha_i y_i + (1 - \alpha_i) x_i$ とする。 x_i と y_i はそれぞれ DCN (Death Certificate Notification) 割合と MI 比 (Mortality Ratio/Incidence rate) と呼ばれ、既知定数である。DCN 割合は罹患数のうち死亡票で初めて確認された割合、MI 比はがんの死亡数を罹患数で割ったものであり、どちらも登録の完全性を表す指標として用いられてきたものである。

$\alpha_i \in (0, 1)$ は x_i と y_i の内点を表現するパラメータである。手法1では各地域独立、手法2では混合効果モデルを仮定して事前分布を置き、信頼区間の構成を行う。具体的には手法1では全ての α_i に独立に一様分布の事前分布を与える。手法2では $\eta_i = \log(1/\alpha_i - 1)$ に事前分布 $N(\mu, \sigma^2)$ を置く。加えて $\pi(\mu, \sigma^2)$ には客観事前分布を与える。

当日はシミュレーションと実データ解析の結果を紹介する。シミュレーションでは、真の罹患数への一致性を持っているかを確認、実データ解析では都道府県ごとのデータを用いて、都道府県ごとの罹患数の信頼区間の推定を行い、また事後分布の形状に大きく影響を与える要因を考察する。

3. 参考文献

[1] Kamo, K. I., Kaneko, S., Satoh, K., Yanagihara, H., Mizuno, S., & Sobue, T. (2007). A mathematical estimation of true cancer incidence using data from population-based cancer registries. Japanese journal of clinical oncology, 37(2), 150-155.