

厚生労働科学研究費補助金 (がん対策推進総合研究事業 (がん政策研究事業))
分担研究報告書

がん登録データの統計モデリング構築及びシミュレーションシステム整備

分担研究者 加茂憲一 札幌医科大学 医療人育成センター 准教授
研究協力者 伊森晋平 大阪大学大学院 システム創生専攻 助教
研究協力者 田辺竜ノ介 大阪大学大学院 システム創生専攻 D3
研究協力者 福井敬祐 大阪国際がんセンター 疫学統計部 研究員

研究要旨

都道府県規模で集約されるがん罹患数について、報告値の分布や信頼性を評価するための数理モデルを混合効果モデルに基づいて構築し、MCMCに基づく罹患数の分布および信頼区間を算出した。具体的には、都道府県規模での罹患数に関する登録完全性についてはロジスティック回帰モデルによる補正を行い、同時に地域依存の全国からのばらつきに関する部分は混合効果モデルを適用し、この両方により頻度論に基づく真の罹患数の分布を推定した。実際に2013年の34府県のデータを用いて各府県の罹患数の分布を推定した。

A. 研究目的

都道府県規模で集約される「がん罹患数」について、統計モデルに基づく推論を与えることが本研究の目的である。罹患数を考察する際には、登録の完全性に関する問題点を無視することはできない。これは、DCNの存在により示唆される問題点である。つまり、DCNが存在する以上、登録漏れかつ生存しているケースが一定数存在している可能性が否定できず、この要因に関する調整を行うことにより真の罹患数が把握できることが期待される。この点に関しては、登録率を推定する決定論的アプローチや、登録漏れ発生に一定のメカニズムを仮定する確率論的なアプローチが存在し、本研究においては確率論的なアプローチに着目する。具体的には、登録に関する

完全性の指標として活用されているDCN割合とMI比に着目し、これら要因間の関連性に回帰モデルにより表現することにより、真の罹患数を推定するというアプローチを用いる。罹患数に関する点推定ベースの議論に関しては、このようなアプローチにより解決されるが、頻度論に基づく分布まで推定するとなると新たな問題が発生してくる。それは都道府県規模での罹患数は数千～万の規模で観察され、罹患数と密接に関連すると考えられる人口や死亡に関しても同様のオーダーであることに起因する。統計学的アプローチにおいては、一般に大数の法則が成り立つことから、サンプル数が極めて大きい本研究テーマにおいては、推定に関する過剰な安定性が発生する。このことは点推定には影響を与えない

が、区間推定や分布の推定においては、非現実的な結果を招く危険性がある。実際に昨年度の報告書に記したとおり、MI比を罹患から死亡が発生する確率とみなし、比率に関する検定を行ったところ多くの都道府県において有意に期待値と異なるという非現実的な結果が得られた。これは膨大なサンプルを有する罹患数の解析において、地域間のMI比に関するばらつきを偶発変動とみなすことに無理があることを示唆する結果である。そこで昨年度報告書においては混合効果モデルによる解決を試みた。本年度は、MI比-DCN割合に回帰構造を導入することにより登録の完全性の補正を行いつつ、階層ベイズモデルにより地域間のばらつきを評価するモデルを用い、都道府県別罹患数に関する分布をMCMCに基づき推定した結果を報告する。

B. 研究方法

MI比とDCN割合との関連性に関する構造としては

$$\text{MI比} = \alpha + (1-\alpha) \text{DCN割合}$$

を仮定した。ここでパラメータ α は登録の完全性を補正した真のMI比の全国値を意味する。本設定はDCN割合=1においてMI比=1という基本的前提を充たしていることにも注意しておく。さらにこの式の幾何的な意味としては、MI比の報告値が、1とDCN割合を $\alpha : 1-\alpha$ に内分する内点であることを意味しており、これは

$$\text{DCN数} \leq \text{報告罹患数} \leq \text{真の罹患数}$$

と同義である。ここでパラメータ α を全国

共通のものと設定せず、都道府県固有の α_i と仮定し (i は都道府県インデックス)、その分布を

$$\begin{aligned} \eta_i = \text{logit}(\alpha_i) &\sim \text{i.i.d. } N(\mu, \sigma^2) \\ \mu &\sim N(0, \tau^2) \\ \sigma^2 &\sim \text{IG}(c, d) \end{aligned}$$

によって設定することにより、都道府県別罹患数の真値に関する分布を推定する。

C. 研究結果

2013年の全がん・男女計に関する34府県(表1参照)のデータ(罹患数、死亡数、DCN数)を用い、前述の手法に基づく府県別罹患数の推定を行った結果を表1および図1に示す。表1は罹患数の報告値と、点推定値およびその95%信頼区間を示す。図1は、この結果をMI比に換算し、DCN割合との関連性をプロットしたものである(横軸がDCN割合、縦軸がMI比を表す)。「○」が報告値、「△」が点推定値を表し、「△」に被る線分が95%信頼区間を表す。また図1中の直線はMI比とDCN割合の関係性を表す回帰直線であり、この回帰直線の切片は全府県共通の真のMI比を意味する。この全地域共通のMI比を横軸方向に破線で延長している。

表1：府県別罹患数の報告値、推定値および95%信頼区間

府県	罹患数			
	MCIJ 2013	点推定	信頼区間	
			2.5%	97.5%
秋田	9352	10548	9207	9641
山形	9221	11458	10550	11191
福島	13696	19988	15813	16619
茨城	18650	29792	20933	21754
栃木	12901	20096	15688	16576
群馬	13171	20060	14071	14696
千葉	34500	63436	38752	39906
新潟	16952	23513	17275	17884
石川	9070	11733	10407	11126
福井	5371	8004	5502	5880
山梨	5116	8548	5349	5699
長野	15073	21334	16870	17719
静岡	23181	37643	26212	27157
愛知	43444	75443	50536	52047
三重	12047	18555	13648	14382
滋賀	8485	14373	9194	9756
大阪	61259	89418	68671	70261
奈良	10087	13953	11508	12220
和歌山	8007	9870	9098	9724
鳥取	4873	5814	4981	5348
島根	5720	7063	6095	6501
岡山	13582	19494	14886	15617
広島	23953	28687	25811	26870
山口	10749	14337	11832	12416
徳島	5098	7754	5880	6359
香川	7054	9908	7911	8465
愛媛	10444	14171	12101	12839
高知	5404	7488	6095	6559
福岡	34257	51199	39637	40888
佐賀	5601	8448	6002	6394
長崎	10599	14034	10930	11430
熊本	12176	18139	13750	14439
大分	7887	11948	9219	9827
沖縄	7363	14221	7893	8409

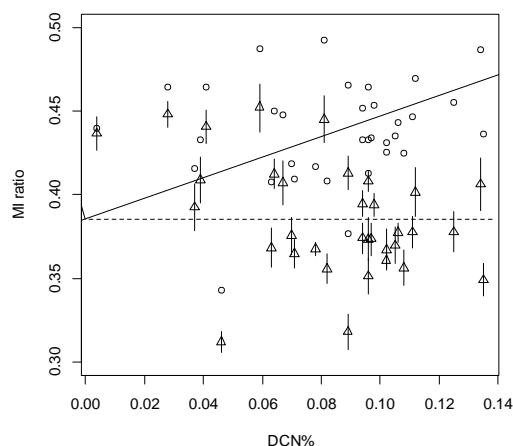


図1：MI比の区間推定

D. 考察

府県別の罹患数について、登録の完全性の指標であるMI比とDCNとの関係性を定式化することにより登録の完全性を補正しつつ、府県間のばらつきを階層ベイズモデルにより調整して評価する手法を確立した。その結果をMI比として表現したのが図1である。図1の実測MI比(「○」のプロット)には右上がりの傾向が見られ、これはDCN割合が低い地域のMI比が低くなる(死亡に対する罹患割合が高くなる)ことを意味する。一方で、真のMI比(「△」のプロット)は横ばいの傾向が見られ、これは真のMI比はDCN割合の影響を受けないことを意味する。真のMI比は登録の完全性に関する補正が行われていることから、これらの傾向は妥当なものであると考えられる。

次に各府県において推定された結果に着目するが、34府県すべての真の罹患数の分布を列挙するには紙面スペースの問題があるため、DCN割合とMI比に関して特徴的

様式A (8)

な特性（最大と最小値）を有する4県の結果のみを図2に示す。図2は順に秋田県（最小DCN割合）、栃木県（最大DCN割合）、広島県（最小MI比）、佐賀県（最大MI比）における真の罹患数の分布の推定結果である。ここで、横軸は罹患数、縦軸は頻度を表す。また、実線が実測の罹患数、ヒストグラ

ムが真の罹患数の分布を推定した結果を表す。基本的には一峰で左右対称に近い形状の分布が推定された。また、真の罹患数は登録の完全性に関する補正がなされるため、分布の中央は実測値よりも高い値となる傾向にあった。

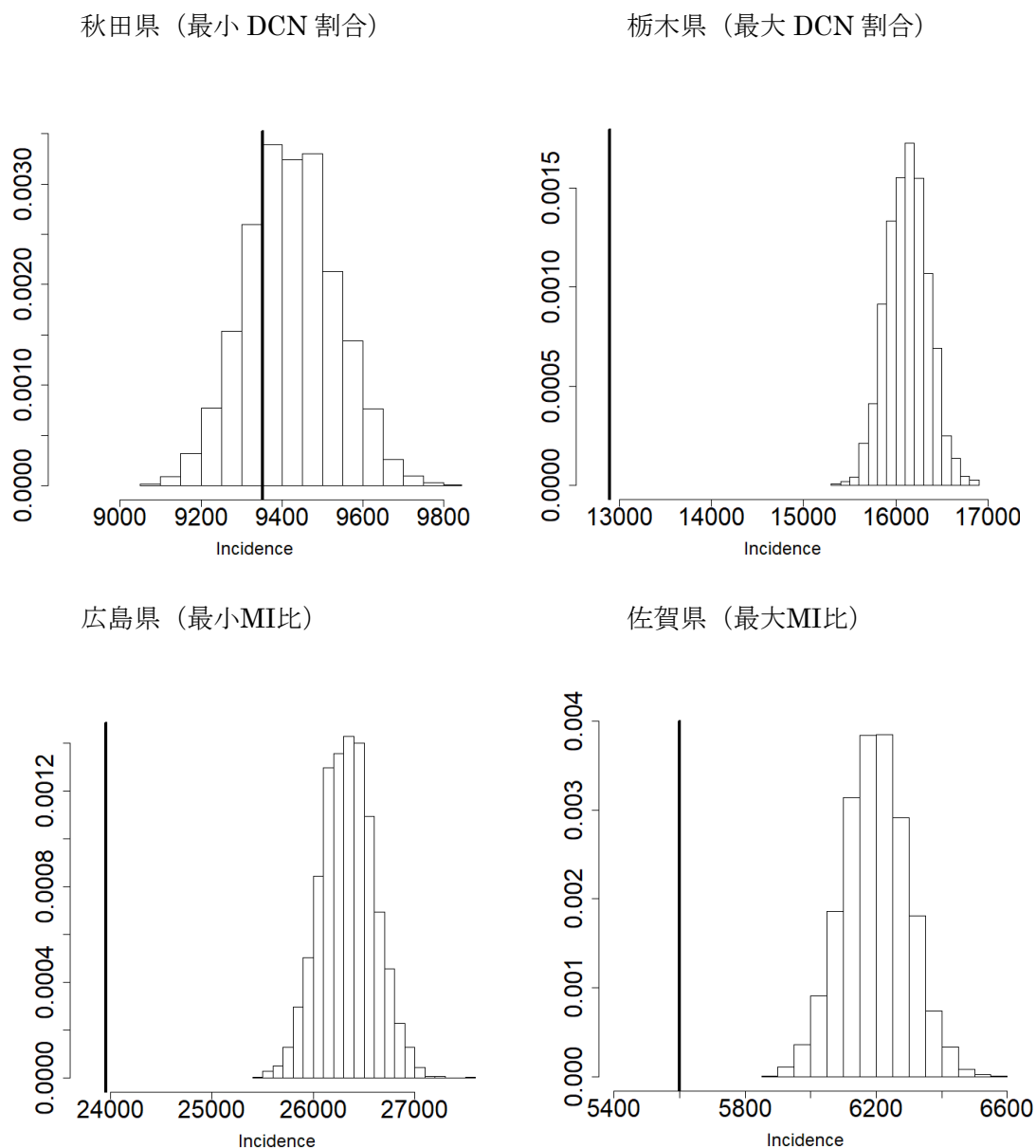


図2：代表的な特徴を有する4県における罹患数分布の推定結果

E. 結論

都道府県における真の罹患数について統計学的推測を与えるために、点推定に関しては登録の完全性を表す指標を用いた回帰構造による補正を行い、分布推定に関しては地域依存のばらつきを考慮に入れた階層ベイズモデルを構築した。実測値はDCN割合の影響を受ける傾向にあったが、推定結果はこの点が補正されていることが確認された。信頼区間に関しては、単純な比率の区間推定による結果に比べて広めの区間が推定され、現実的な結果であった。実際の分布の形状も一峰の対称形であり、特異な形状ではなかった。県別の結果を確認すると、DCN割合の低い秋田県においては、実測の罹患数と、補正された罹患数が近い数値であり、実測値が95%信頼区間に含まれることから、低DCN割合地域における報告値の信頼性が高いことが窺える。一方で信頼区間の幅に影響を与える要因を探索したが、候補として設定した「人口」「DCN割合」「期待値と実測値の乖離」について際立った傾向は観察されなかった。

F. 健康危険情報

特になし。

G. 研究発表

1. 論文発表

- 1) K.Kamo, T.Tonda, K.Satoh: Growth analysis using nuisance baseline. FORMATH, 16, 1-10, 2017.
- 2) H.Yanagihara, K.Kamo, S.Imori, M.Yamamura: A study on the bias-correction effect of the AIC for

selecting variables in normal multivariate linear regression models under model misspecification. REVSTAT-Statistical Journal, 15 (3), 299-332, 2017.

- 3) A.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 (accepted).
 - 4) 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 (accepted).
 - 5) 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 ^{125}I iodine source during implantation in brachytherapy. Journal of Radiation Research (accepted).
- ### 2. 学会発表
- 1) 加茂憲一, 田辺竜ノ介, 伊森晋平, 福井敬祐: 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. 日本

- がん登録協議会学術総会(愛媛県医師会館), 2017年6月8~10日, ポスター.
- 2) 加茂憲一, 伊藤ゆり, 福井敬祐, 片野田耕太: シミュレーションモデルを用いた大腸がん死亡リスク低減の定量化. がん予防学術大会(愛媛県医師会館), 2017年6月16~17日, 口演.
- 3) 中山富雄, 伊藤ゆり, 福井敬祐, 加茂憲一, 雑賀公美子. がん検診対象者の年齢上限に関する検討. がん予防学術大会(大阪国際がんセンター), 2017年6月16~17日, 口演.
- 4) Kamo, K.Fukui, Y.Ito, K.Katanoda. Microsimulation model for colorectal cancer to estimate effect of FOBT screening programme and improvement in cancer care in Japan: CAMOS-J CRC, The 21st International Epidemiological Association World Congress of Epidemiology (ソニックシティ), 2017年8月19~22日, 口演.
- 5) 田辺竜ノ介, 加茂憲一, 伊森晋平, 福井敬祐. 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, 統計関連学会連合大会(南山大学), 2017年9月3~6日, 口演.
- 6) 富田哲治, 加茂憲一, 佐藤健一. 経時離散データに対するベースラインを特定しない変化係数の推測について, 統計関連学会連合大会(南山大学), 2017年9月3~6日, 口演.
- 7) 福井敬祐, 伊藤ゆり, 加茂憲一, 片野田耕太, 中山富雄: マイクロシミュレーションを用いた大腸がん検診による死亡率減少効果の推定, 日本疫学会学術総会(コラッセ福島), 2018年2月1-3日, 口演.
- 8) 富田哲治, 加茂憲一, 佐藤健一. 局外ベースラインのある時点間相関を考慮した回帰モデルによるがん統計データの分析, 日本疫学会学術総会(コラッセ福島), 2018年2月1-3日, ポスター.

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