

学術奨励賞受賞講演**地域がん登録データの時系列解析
Trend analysis of population-based cancer registry data**

独立行政法人国立がん研究センターがん対策情報センターがん統計研究部 片野田耕太

1. はじめに

地域がん登録のデータを用いた3つの研究について紹介したい。いずれの研究も、地域がん登録データの時系列に関するものである。罹患の年次推移は、がんの統計において長年懸案になっていた。国立がん研究センター、古くは大阪府立成人病センターが罹患の全国推計値を1975年以降の値として公表してきたが、登録精度の向上や推計に含まれる地域の出入りの影響が含まれることから、その解釈が困難であった。そこで、長期間にわたって精度が安定した地域に絞って年次推移の検討ができないか、というのが1つ目の研究の出発点である。

日本では死亡データを使ってがんの長期トレンドを見ることができる。死亡データで見ると、主要ながんの年齢調整死亡率は1990年代半ばで増加が収束している。これが予防危険因子の影響なのかを知るためには、罹患でも同じようなトレンドが見られるかを調べなければならぬ。これが2つ目の研究の着眼点であった。

がん罹患データの最新年は、死亡データの最新年より4~5年遅れている。標準化および精度向上の努力によりこの遅れは1年ほど短縮されたが、2014年時点で罹患の最新年は2010年である。罹患データの集計に時間がかかるのはデータの性質上仕方がなく、国際的に見ても4年程度の遅れは標準的である。そこで、数理的な手法でこの遅れを埋めることはできないか。これが3つ目の研究の発想である。特に、出生年による罹患率、死亡率の違いが顕著な男性肝臓がんにおいて、1990年代半ばに生じた増加の収束を数理的に予測できるかどうか重要な点であった。

2. 方法**(1) 高精度地域のデータを用いた罹患の年次推移の検討¹⁾****1) データ**

宮城、山形、福井、大阪、および長崎の5府県の1985~2004年がん罹患データ

を用いた。これらの府県は、1980年代から登録精度が安定していると考えられることから候補地域として選ばれた。

全国についての代表性を確認するために、1995～2004年のがん死亡データを人口動態統計の公表値から得た。

2) 統計解析

代表性の確認は、1995～2004年の10年間の年齢調整罹患率および年齢調整死亡率に対数線形モデルを適用し、その回帰係数について funnel plot を行うことにより行った。これは、推定された回帰係数の値と、その標準誤差の逆数との関係をプロットしたものである。

(2) 罹患と死亡の年次推移の分析²⁾

1) データ

がん死亡データについては1958～2011年の人口動態統計より（結果には1970年以降を示す）、がん罹患データについては1985～2007年の宮城、山形、福井、および長崎県の地域がん登録より得た。これらの4県は、(1)の研究で日本全体への代表性が確認されたことから選ばれた。

2) 統計解析

年齢調整死亡率、および年齢調整罹患率の時系列データに対して、Joinpoint 回帰分析と呼ばれる折れ線の当てはめを行った。この手法は、米国 National Cancer Institute が開発したもので、がんの増減の判定を行う方法として国際的に広く用いられている。

(3) 罹患データの短期予測³⁾

1) データ

1985～2007年の宮城、山形、福井、および長崎県の地域がん登録の罹患データ((2)と同じ)。

2) モデル

年齢、暦年にスプライン関数を当てはめてそれらのその交互作用を含めて説明変数とした一般化加法モデルを用いた(A*Pスプライン)⁴⁾。モデルの比較対象としては、年齢および暦年の線形モデル(A+P1次)と、年齢、暦年、およびそれらの交互作用の線形モデル(A*P1次)を用いた。さらに、年齢、暦年、およびそれらの交互作用の2次モデル(A*P2次)も追加で検討した。

3) 統計解析

男性肝臓がんは、1990年代半ばに罹患数の増加が収束していることから、1985

～1995年までのデータを用いて2000年の値を予測し、その精度を調べた。本稿では男性肝臓がんの結果のみを示す。他の結果については引用文献3を参照されたい。

予測精度は、予測値の信頼区間が実測値を含んでいるか、および予測誤差（予測値/実測値-1）で評価した。

3. 結果

(1) 高精度地域のデータを用いた罹患の年次推移の検討

図1に全がん死亡の funnel plot を示す。男性、女性とも、大阪府を除く4県は、回帰係数（年齢調整死亡率の変化率に当たる）が全国値の95%信頼区間内にあったが、大阪府だけは全国値の95%信頼区間よりも小さな値、つまり減少率が全国よりも急であった。女性では、5府県の合計値も全国値の95%信頼区間から外れていた。4県の合計値は、男女とも全国値の95%信頼区間内にあった。

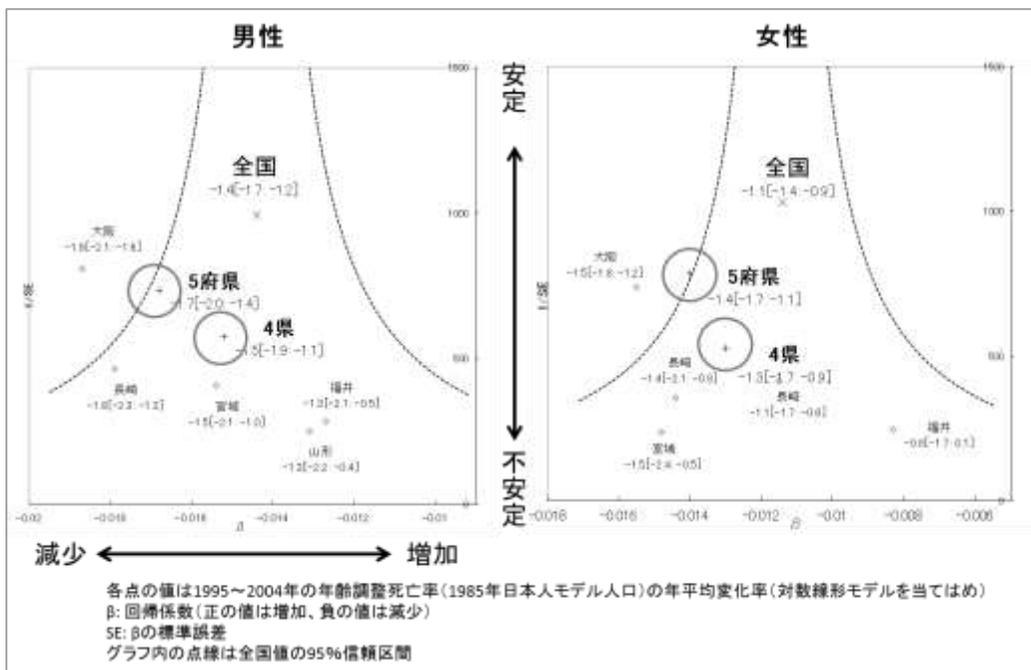


図1 全がん死亡の funnel plot

(2) 罹患と死亡の年次推移の分析

図2に全がん罹患および死亡の年齢調整率の年次推移の結果を示す。罹患では、男女とも1985年以降単調かつ有意な増加であった（年変化率：男性0.6% [95%信頼区間0.5: 0.8]、女性0.8 [95%信頼区間0.7: 0.9]）。しかし、前立腺がんを除くと男性の罹患の増加は1992年までで、その後統計学的に有意な減少に転じた（年変化率-0.3% [95%信頼区間-0.5: -0.2]）⁵⁾。

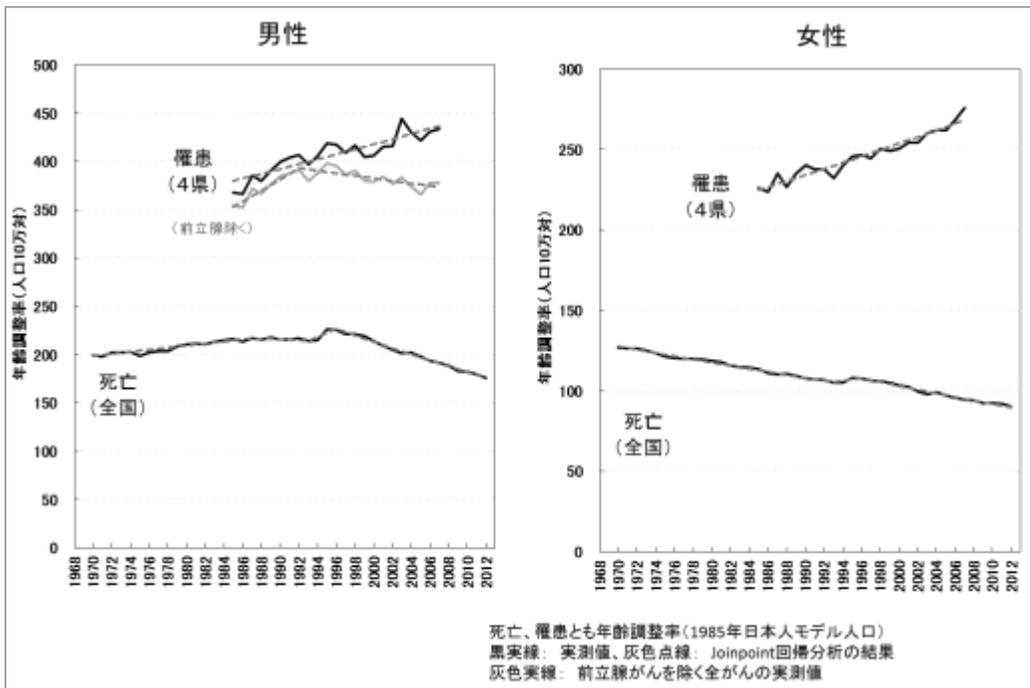


図2 全がん罹患および死亡の年次推移

(3) 罹患データの短期予測

図3に男性肝臓がんの予測結果を示す。A*P スプラインモデルが1990年代半ばの罹患数の増加収束を予測できたのに対して、他のモデルは1990年代半ば以降も罹患数が増加し続ける予測結果であった。また、A*P スプラインモデルの予測値の95%信頼区間は、予測期間の実測値をすべて含んでいた。

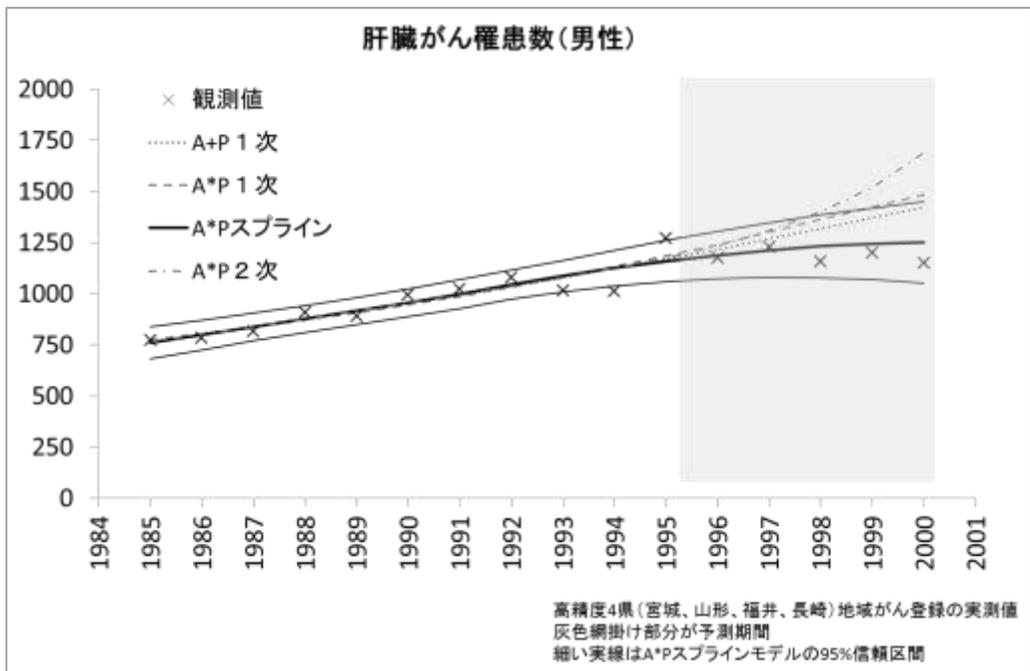


図3 男性肝臓がんの短期予測結果

4. 考察

1つ目の研究は、大阪府を含んだ5府県のデータよりも、大阪府を除いた4県のデータの方が日本全体のがんの年次推移の代表性が高いことを示した。人口規模的には、これら5府県のうち大阪府が半分近くを占めるため、罹患数の安定性をとるか、年次推移の代表性をとるかのトレードオフであった。本稿には含めなかったが、罹患で5府県と4県の年次推移の傾向を同様に検討すると、5府県では直近の数年での罹患率減少ががん種共通で見られ、罹患の年次推移においても大阪府が他県と異なる傾向を示すことが示唆された(1)。これらを総合的に判断して、4県のデータを罹患の年次推移を調べる標準的な手法として採用することとした。現在では、これら4県のデータが国立がん研究センターがん対策情報センターのウェブサイト「がん情報サービス」で閲覧、ダウンロード可能である(<http://ganjoho.jp/public/statistics/pub/statistics02.html>)。

2つ目の研究は、1つ目の研究結果を受けて、日本のがん死亡とがん罹患の年次推移を総合評価した試みである。全がんでは男女とも年齢調整罹患率が増加傾向にあったが、男性では前立腺がんを除くと1990年代で減少に転じていた。がん種別に詳しく見ると、大腸がん、

肝臓がん、男性肺がんで 1990 年代半ばに罹患の増加が止まっており、死亡のトレンドと類似していた 2)。理論的には罹患の減少が死亡の減少に先行しているべきではあるが、人口集団全体のデータで、罹患と死亡がほぼ同じ 1990 年代に増加が収束したことは、予防危険因子の影響であることを強く示唆する。

3 つ目の研究は、年齢、暦年、およびその交互作用によるモデルが (A*P スプライン)、日本の肝臓がんのように、生まれた年でリスクが異なるがんの増減を高精度で予測できることを示した。本稿には含めなかったが、がん種別の結果を見ても、子宮頸がんと前立腺がんを除けば、±10%程度の誤差での予測が実現できていた 3)。子宮頸がんと前立腺がんは、2001 年から 2005 年の間に急激な罹患の増加が観察されており、1985~2000 年のデータを用いた 2005 年の予測において、A*P スプラインモデルでも 30%程度の過小評価であった。増加が収束した 2005 年までのデータを用いて 2007 年の予測を行った追加解析では、子宮頸がんについては予測誤差が 4%程度に縮小したが、前立腺がんについては 17%程度の過大評価となった 3)。前立腺がんについては PSA 検診の普及状況で罹患の増減が大きく変わる可能性があり、過去のデータを外挿する本研究のような予測モデルでは限界がある。

日本の肝臓がんは、罹患率、死亡率とも 1930 年代生まれに急峻なピークがあり、出生年を説明変数とした予測モデルの精度を調べる試金石として有用である。予備解析の段階で、国際的によく用いられている Nordpred[®]を試したが、少なくとも 5 年間の予測においては、1990 年代半ばの増加収束は予測できなかった。Nordpred が 5 年平均のデータを用いていること、直近のトレンドをモデルに組み入れていること (4 県データでは 1995 年に一時的に男性肝臓がん罹患数が増加している) などが影響している可能性がある。

短期予測は、長期予測と異なり数年経つと実測値との比較ができる点が厳しいところである。米国では、短期予測の方法が数年おきに変更されている 7)。本研究で採用した手法も、今後の実測値との比較の結果次第で修正する必要性が生じるかもしれない。

予測精度の限界はあるにせよ、A*P スプラインモデルを用いることで日本の罹患データの最新値を一定の精度で得ることが可能となった。この手法は死亡データにも適用可能なので、米国のようにリアルタイムの罹患数と死亡数の推計が可能である。「がん情報サービス」では、2014 年のがん統計予測としてそれらの結果を掲載している

(http://ganjoho.jp/public/statistics/pub/short_pred.html)。

3 つの研究を通して、罹患と死亡を組合わせた年次推移の検討と、リアルタイムの罹患数と死亡数の推計が可能となった。これらのデータをわが国のがん対策の立案と評価に活用し、がんという疾病のコントロールにつなげることが今後の課題である。

5. 参考文献

1. Katanoda K, Ajiki W, Matsuda T, et al. Trend analysis of cancer incidence in Japan using data from selected population-based cancer registries. *Cancer Sci.* Feb 2012;103:360-368.
2. Katanoda K, Matsuda T, Matsuda A, et al. An updated report of the trends in cancer incidence and mortality in Japan. *Jpn J Clin Oncol.* May 2013;43:492-507.
3. Katanoda K, Kamo K, Saika K, et al. Short-term projection of cancer incidence in Japan using an age-period interaction model with spline smoothing. *Jpn J Clin Oncol.* Jan 2014;44:36-41.
4. Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics.* Oct 2005;6:576-589.
5. 片野田耕太, 松田智大, 松田彩子, et al. 地域がん登録データを用いたがん罹患の長期トレンドの分析. *JACR Monograph.* 2013(19):153-159.
6. Nordpred software package.
<http://www.kreftregisteret.no/en/Research/Projects/Nordpred/>. Accessed Aug. 8, 2012.
7. 片野田耕太. 米国のがん統計に用いられている数理モデルの概観. *統計数理.* 2011;59:173-180.

Summary

Trend analysis of population-based cancer registry data is essential for cancer control. Three studies are summarized in this article: (1) trend analysis of cancer incidence in Japan using data from selected population-based cancer registries, (2) An updated report of the trends in cancer incidence and mortality in Japan, and (3) Short-term projection of cancer incidence in Japan using an age-period interaction model with spline smoothing. The first study revealed a great geographical variation in the trend of cancer incidence in Japan. In particular, Osaka prefecture showed a steeper decrease than four other prefectures (Miyagi, Yamagata, Fukui, and Nagasaki) in both cancer incidence and mortality. The second study showed that age-standardized incidence rate for major cancers increased until 1990s and decreased or stabilized thereafter, including colorectal, liver, and male lung cancer. The third study examined the predictive performance for short-term projection of several mathematical models. A model using age, period, and their interactions with spline smoothing was shown to have a good predictive performance for a cancer with a strong cohort effect, such as liver cancer. These statistical tools should be used to plan and evaluate cancer control policies in Japan.