

サポートベクター回帰による 都道府県別乳がん死亡率の要因分析

鈴木孝弘^a・田辺和俊^b

Analysis of Factors Related to Prefectural Mortality Rates of Female
Breast Cancer by Using a Support Vector Regression

Takahiro SUZUKI^a・Kazutoshi TANABE^b

Abstract

Breast cancer is one of serious social issues especially for women in Japan. A statistical analysis by using a multivariate nonlinear regression method has been tried to reveal factors related to prefectural mortality rates of female breast cancer from the viewpoint of regional countermeasure to the falling birthrate. Age-adjusted mortality rates of female breast cancer for 47 prefectures were used as an objective variable, and 40 indices in fields of lifestyle habit, food intake and socio-economics as explanatory variables. The factors related to the mortality rates were examined by applying a support vector regression method to these data, and their sensitivities to the mortality rates were estimated by employing the sensitivity analysis method. Thirteen kinds of factors which satisfactorily reproduce the observed mortalities of 47 prefectures with high accuracy were obtained. Among them, smoking and obesity highly contribute to the observed mortality rates while part of factors such as alcoholic drinking, intake of milk and dairy products and of soy products show tendencies to the mortality different from the results so far of epidemiological studies. For Aomori prefecture where the mortality rate of female breast cancer is the highest in Japan, regional countermeasures to the falling birthrate were examined, and several recommendations were attempted.

Keywords : Breast cancer, Prefectural mortality rates, Analysis of related factors, Support vector regression method

^a 東洋大学自然科学研究室：〒112-8606 東京都文京区白山 5-28-20

Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan

^b 東洋大学現代社会総合研究所：〒112-8606 東京都文京区白山 5-28-20

Institute of Social Sciences, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan

1. はじめに

欧米先進国と同様に日本でも乳がんの罹患率（1年間に新たに発見された患者の割合）や死亡率が増加しており、乳がんの予防対策が重要となっている。1980年代から現在までの年齢調整罹患率の推移を見ると、男女の胃、肺、大腸のがんが微減または停滞傾向を示しているのに対し、女性の乳がんは1985年の28.5（人口10万人対）から2012年には70.0へ2.5倍も増加した。また、1998年に胃がんを抜いて女性のがん1位となり、2012年には女性の全部位がんの約24%を占めるまでになった。一方、乳がんは他部位のがんと比べて早期発見が可能であり、予後が良好で致命率が低いことから、死亡率は罹患率よりかなり低い。しかし、全部位がんの人口調整死亡率が1950年代から激減する中で、乳がんは罹患率の増加のために人口調整死亡率も1990年の8.2（人口10万人対）から2015年には12.0まで増加した。さらに、2008年には胃がんと肺がんを抜いて女性がん1位となり、2015年には全部位がん中の約14%を占めている。

日本人女性の乳がんの増加は生活スタイルの欧米化が影響し、乳がんのリスク要因として生活習慣が重要であるとされている。これまで欧米諸国を中心にがんと生活習慣や食物等の関連を科学的に検証した多くの疫学研究が行われてきている。World Cancer Research FundとAmerican Institute for Cancer Researchは種々のがんと生活習慣等の関連についての多数の論文を評価し、報告書“Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective”¹⁾を発表した。この中で乳がんについては、危険要因として飲酒、肥満等、予防要因として運動等をあげている。また、日本人の乳がんについては幾つかの評価報告があり、危険要因として肥満、飲酒、喫煙等、予防要因として運動、授乳等があげられている。

しかし、乳がんについてはこれら以外に様々な要因が上げられており、乳がん予防対策の関係者にとっては明確な予防指針が下せる状況ではない。予防対策のためには、乳がんの発症に重大な影響を与える影響要因の種類とその乳がん発症に対する影響度を明らかにする必要がある。一方、国内においては地域により、乳がんの罹患率・死亡率にはかなり差があることが認められており、地域要因の解明にも関心が集まっている。そこで、地域の罹患率または死亡率のデータを目的変数、幾つかの指標を説明変数として重回帰分析する生態学的分析手法が注目され、この手法によって都道府県の乳がん死亡率の解析から影響要因を探索した幾つかの研究がある。しかし、これまでの研究では特定分野の比較的少数の説明変数を用いて線形回帰分析（OLS）により解析しているため、統計的に有意な回帰結果を得た研究はなく、乳がん発症の影響要因の種類とその影響度は未だ明らかになっていない。

本研究では地域の乳がん対策に資する情報提供の可能性を探る観点から、都道府県別の女性乳がん罹患率または死亡率を対象として、非線形重回帰分析により統計的に有意な要因を見出すことを目的とする実証研究を試みた。

2. データと方法

2.1 都道府県別の女性乳がんのデータ

女性乳がんに関する都道府県別データとしては、罹患率、患者率（受療中の乳がん患者の割合）、死亡率の3種のデータが政府統計として公開されている。この中では罹患率または患者率のデータを解析してその要因を解明することが乳がん予防の観点からは有効と考えられる。しかし、罹患率については、国立がん研究センターから「がん情報サービス」により都道府県別の数値が公開されているが、都道府県により乳がんの判定水準が異なり、罹患者の登録精度に差があるなどの理由から、乳がんの要因分析に採用することには問題がある。また、患者率については、厚生労働省の「患者調査」により各都道府県の患者数が公表されている。しかし、データが千人単位のため、患者の少ない都道府県では公表データが1桁の数値、または数値0になる地域があり、正確な患者率の推計が困難である。

そこで、目的変数には厚生労働省の「人口動態調査特殊報告」による2015年の都道府県別年齢調整死亡率（女性）を用いた。そのデータは図1に示すように青森県の死亡率が15.6（女性人口10万人当たり）と突出して高く、最少の鳥根県の9.5との差は大きい。地域別の傾向では山形県を除く関東地方以北が全般的に高いが、中部地方以西は一部の府県を除いて全般的に低い。また、鳥根と山口、福岡と大分のように死亡率が高い府県と低い府県が隣接している地域もある。このように複雑な地域格差は多種多様な要因が死亡率に関与している可能性を示唆する。

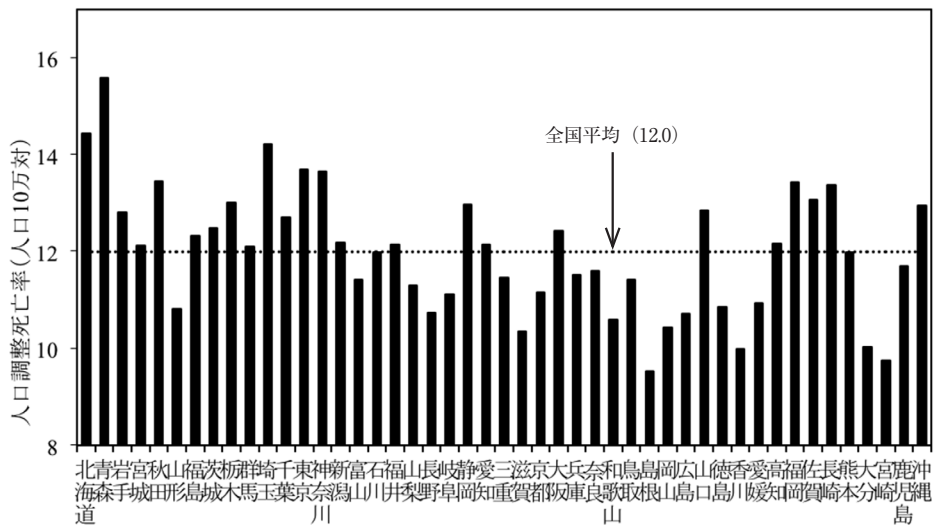


図1 都道府県別の女性乳がん死亡率

2.2 説明変数

重回帰分析の説明変数としては、まず、これまでの疫学的研究で検証されている多数の乳がん要因をできるだけ採用した。さらに、乳がんの罹患や死亡への影響が報告されているが、これまで回帰分析の説明変数として検証されていない数種の新規指標があり、それらを選定した。

乳がんに関する多数の疫学論文を評価した研究としては、前記のWorld Cancer Research Fund (WCRF) とAmerican Institute for Cancer Research (AICR) は“Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective”の中で、乳がんに対して飲酒、肥満、運動等の要因の確実度をConvincing (確実), Probable (ほぼ確実), Limited-suggestive (可能性あり), Limited-no conclusion (証拠不十分), Substantial effect on risk unlikely (大きな関連なし)の5段階で評価している。また、International Agency for Research on Cancer (IARC) は“IARC Monographs on the Evaluation of Carcinogenic Risks to Humans”²⁾の中で各種の発ガン物質の評価を行っている。

一方、わが国では、1986年から1988年にかけて乳癌研究会の「乳癌のhigh risk groupとは」研究班がまとめた乳がんのリスクファクター評価結果³⁾があり、肥満、飲酒以外に高齢、未婚等の危険度も評価している。また、日本乳癌学会は2004年から発表している「科学的根拠に基づく乳癌診療ガイドライン②疫学・診断編」⁴⁾の中で、WCRF/AICRやIARCの結果を参考に日本人についての論文評価も行い、生活習慣要因と乳がんの関連度を5段階で評価している。さらに、国立がん研究センターの「科学的根拠に基づく発がん性・がん予防効果の評価とがん予防ガイドライン提言に関する研究」班では、2003年から日本人を対象とした研究に基づいて生活習慣や生活環境などの危険度を4段階で評価した結果⁵⁾を公表している。

以上の5団体の評価を比較して表1に示すが、評価が一致していない要因が多い。また、閉経前と閉経後では評価が異なる要因もあるが、これは乳がんの発症機構が閉経前と閉経後では異なるため、乳がんリスクにも違いがあると考えられているからである。本研究ではこれらの評価結果等を参考に、乳がんの死亡に影響すると考えられる要因を探索した結果、採用した説明変数の総数は表2に示す40種となった。この表で8種類の太字の変数はこれまで未検証の新規変数である。なお、これら説明変数候補のデータは、現時点で入手可能な最新のものであり、乳がんの発症時のものではなく、また医学的にはがん細胞が生じたときとの時間的ずれがある。そのため、本研究では、これらの説明変数の経年変化は無視できると仮定している。また、本研究で都道府県別の指標データを取得した各種の政府統計では閉経前後の記載はないため、全年齢での平均に相当するデータであると考えられる。各指標の単位が異なり、また下記の感度分析のために、全指標は最小値と最大値が0と1になるよう正規化して解析に用いた。

表1 国内外の諸団体による乳がん要因に関する評価結果

要因	WCRF-AICR (2018)	IARC (2012)	乳癌研究会 (1988)	日本乳癌学会 (2018)	国立がん研究センター (2018)
成人期の身長	+++				
出生時体重	++/?				
肥満	- - / +++		+	+/+++	+/+++
運動	- / - -			? / - -	-
喫煙		+		++	+
受動喫煙				+	+
飲酒	++/+++	+++	+	+/++	?
肉・魚・卵	?				?
脂質	?/+			?	?
牛乳・乳製品	- / ?			-	?
野菜・果物・茶	?				?
大豆製品	?			-	-
高齢			+++		
大都市圏			+		
専門・管理職			+		
心理ストレス				?	
未婚			++		
出産歴				- -	
初産年齢高			++	+++	
授乳歴			-	- -	-
初経年齢低	- -		+	++	
閉経年齢高			+	++	

+++：リスク確実，++：リスクほぼ確実，+：リスク可能性あり，-：予防可能性あり，- -：予防ほぼ確実，- - -：予防確実，?：証拠不十分，/のある要因：閉経前の評価/閉経後の評価

表2 要因分析に用いた説明変数

分野	説明変数	定義	データ出所	
生活習慣	喫煙	たばこを毎日または時々吸う女性の比率 (%)	国民生活基礎調査	
	受動喫煙	たばこを毎日または時々吸う男性の比率 (%)	国民生活基礎調査	
	飲酒	酒を飲む女性の比率 (%)	国民生活基礎調査	
	規則食事	規則正しく食事をとっている女性の比率 (%)	国民生活基礎調査	
	バランス食事	バランスの取れた食事をしている女性の比率 (%)	国民生活基礎調査	
	睡眠	睡眠を十分にとっている人の比率 (%)	国民生活基礎調査	
	ストレス	ストレスを抱えている女性の比率 (%)	国民生活基礎調査	
	健康無関心	健康のために特に何もしていない女性の比率 (%)	国民生活基礎調査	
	エネルギー	女性1人1日当たりのエネルギー摂取量 (Kcal)	国民健康栄養調査	
	肥満	女性のBody Mass Indexの平均 (Kg/m ²)	国民健康栄養調査	
	運動	適度な運動や身体活動をしている女性の比率 (%)	国民生活基礎調査	
	交際費	世帯当たりの交際費の年間支出額 (千円)	家計調査	
	摂取食品	炭水化物	女性1人1日当たりの炭水化物摂取量 (g)	国民健康栄養調査
米類		女性1人1日当たりの米類摂取量 (g)	国民健康栄養調査	
そば・うどん		世帯当たりのそば・うどんの年間支出額 (千円)	家計調査	
蛋白質		女性1人1日当たりの蛋白摂取量 (g)	国民健康栄養調査	
動物性蛋白質		女性1人1日当たりの動物性蛋白質摂取量 (g)	国民健康栄養調査	
肉類		女性1人1日当たりの肉類摂取量 (g)	国民健康栄養調査	
加工肉		女性1人1日当たりの加工肉摂取量 (g)	国民健康栄養調査	
魚介類		女性1人1日当たりの魚介類摂取量 (g)	国民健康栄養調査	
卵		女性1人1日当たりの卵摂取量 (g)	国民健康栄養調査	
大豆製品		女性1人1日当たりの大豆製品摂取量 (g)	国民健康栄養調査	
脂質		女性1人1日当たりの脂質摂取量 (g)	国民健康栄養調査	
動物性脂質		女性1人1日当たりの動物性脂質摂取量 (g)	国民健康栄養調査	
牛乳・乳製品		女性1人1日当たりの牛乳・乳製品摂取量 (g)	国民健康栄養調査	
野菜		女性1人1日当たりの野菜摂取量 (g)	国民健康栄養調査	
緑黄野菜		女性1人1日当たりの緑黄野菜摂取量 (g)	国民健康栄養調査	
果物		女性1人1日当たりの果物摂取量 (g)	国民健康栄養調査	
砂糖		女性1人1日当たりの砂糖摂取量 (g)	国民健康栄養調査	
コーヒー		女性1人1日当たりのコーヒー摂取量 (g)	国民健康栄養調査	
緑茶		世帯当たりの緑茶の年間支出額 (千円)	家計調査	
社会経済		都市化	人口集中地区人口比率 (%)	社会生活統計指標
		高齢化	女性総人口に占める65歳以上の女性人口の割合 (%)	社会生活統計指標
	未婚率	女性の生涯未婚率 (50歳時の未婚割合) (%)	社会生活統計指標	
	初婚年齢	女性の初婚平均年齢 (歳)	社会生活統計指標	
	出生率	合計特殊出生率	社会生活統計指標	
	乳がん検診	女性の乳がん検診の受診率 (%)	国民生活基礎調査	
	学歴	女性の就学年数 (年)	社会生活統計指標	
	就業率	女性の就業している比率 (%)	社会生活統計指標	
	女性管理職	管理職にある女性の比率 (%)	社会生活統計指標	

太字の説明変数：これまで未検証の新規変数

2.3 解析方法

本研究では、非線形重回帰分析の手法としてサポートベクターマシン (SVM)⁶⁻⁸⁾ を適用した。SVMは説明変数の数値に対してカーネルと呼ぶ非線形関数を用いて学習パターンを別の空間 (超平面) に写像し、その空間で線形回帰を行う。この操作により、説明変数の元の数値での非線形回帰が可能になり、目的変数と説明変数の間の任意の関係に対して高精度の回帰結果が得られる。説明変数間の交絡効果が予想される場合でもSVMではこの効果は自動的に考慮されるため、変数の積の項の追加は不要である。また、OLSと異なり、SVMでは変数間に強い相関がある場合でも解析可能であり、多重共線性問題は生じない。さらに、目的変数の数以上の説明変数を用いる場合、OLSでは主成分分析等の手法を用いて説明変数を減らす操作を行う必要がある。これに対し、SVMではデータ数以上の説明変数を用いた解析が可能である。以上のSVMの多くの利点はカーネル回帰という手法の採用によるものである。

SVMのソフトウェアはLIBSVM Ver. 3.11⁹⁾ の回帰機能 (SVR) を用いた。多数の説明変数の中から影響要因を探索するためにはSVMモデルと説明変数の最適化を行う必要がある。前者については、LIBSVMのSVRの3種のパラメータ (g , c , p) の最適化を交差検証法により行った。後者に関しては、回帰分析では一般に説明変数の中に有効でないものがあると過学習状態に陥り、学習データに対する誤差は減少するが、予測データについての誤差は増大するため、必要最小限の説明変数を抽出する変数選択が必要である。本研究では迅速な変数選択法として感度分析法を採用した。この感度分析法は、目的変数に対する各説明変数の感度を計算し、感度の低い変数を順次削除しながらSVMモデルを最適化し、目的変数の予測値と実測値の平均二乗誤差が最小となる組み合わせを探索する方法である。筆者らはこの感度分析法の有用性を様々な問題において実証している¹⁰⁻¹⁵⁾。

そこで、以下の手順により影響要因の探索を行った。

- ①データセット中の最初の都道府県を予測データ、その他の46都道府県を学習データとし、パラメータ g , c , p をグリッドサーチして最適条件を探し、このモデルに予測データを入力して死亡率の予測値を求める。
- ②2番目以降の都道府県を予測データとして①の操作を繰り返し、47都道府県について死亡率の予測値と実測値のRMSE (平均二乗誤差; Root Mean Square Error) を求める。
- ③感度を求める説明変数は実際の数値に設定し、その他の変数は全都道府県の平均値に設定したデータをモデルに入力して出力値を求め、当該変数の実測値を説明変数、出力値を目的変数とする単回帰分析を行い、回帰直線の傾きをその変数の感度とする。
- ④全説明変数の中で感度の絶対値が最小の変数を順次取り除いて以上の操作を繰り返し、全都道府県についてのRMSEが最小になる説明変数の組を影響要因とする。

3. 結果

以上の方法により、40種の説明変数の中から影響要因を探索した結果、13種を用いた場

合に死亡率の予測値と実測値とのRMSEが最小となり，その前後でRMSEは増加した。13種を用いたときの回帰決定係数 (R^2) は0.796であり，危険率1%で有意と判定される (図2左)。これに対して，同じデータを用いて線形回帰分析 (OLS) を行ったところ， R^2 は0.553と低くなり，有意とは判定されない (図2右)。この結果から死亡率に対して非線形性関係にある変数が多数あることが示唆され，SVM適用の必要性が実証された。

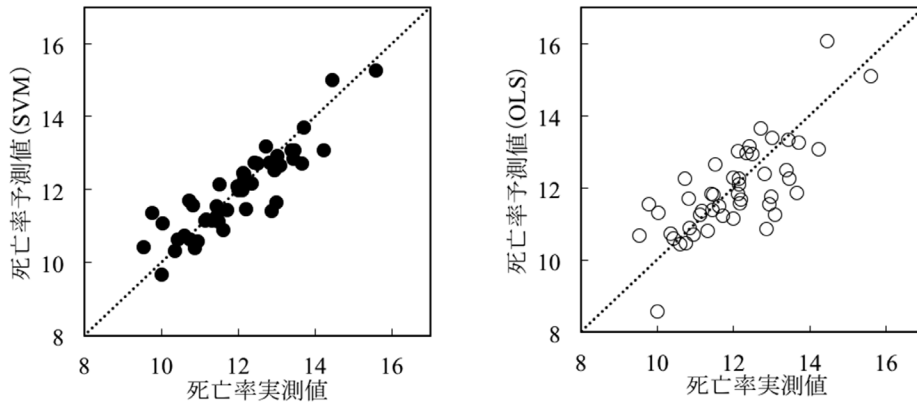


図2 SVM (左) とOLS (右) による死亡率の予測値vs実測値の散布図

得られた各要因の死亡率への相対的な影響度について考察するために，影響要因*i*の感度 S_i から式

$$C_i(\%) = \frac{S_i^2}{\sum_{i=1}^{13} S_i^2} \times 100 \quad (1)$$

により死亡率に対する寄与率 C_i を推定した。影響要因の内訳，死亡率に対する感度とその推定誤差，寄与率，および記述統計を表3に示す。

感度分析で得られた各要因の感度は，他の変数は固定し，当該変数のみ数値を変化させた時の死亡率の変化から求めたので，死亡率に対する当該要因の正味の影響度を表わしている。したがって，感度が正の8種の要因は死亡率増加に寄与する危険要因であり，感度が負の5種の要因は死亡率減少に寄与する予防要因であると解釈できるが，これらの結果はそれぞれの死亡率に対する影響の予測方向と整合する。また，得られた影響要因の感度の誤差は感度分析過程における感度値のばらつきから推定したが，その大きさを見ると，感度はいずれも信頼できる精度で推定されていると考えられる。

表3 影響要因の内訳、乳がん死亡率に対する感度と寄与率、および記述統計

順位	影響要因	感度		寄与率 (%)	記述統計		
		予防要因	危険要因 (推定誤差)		平均	相関係数	
1	喫煙		0.279	0.034	30.0	9.6	0.721
2	肥満		0.173	0.032	11.5	22.4	0.162
3	乳がん検診	-0.156		0.022	9.4	35.3	-0.181
4	運動	-0.151		0.029	8.9	32.4	-0.059
5	牛乳・乳製品		0.138	0.028	7.4	114	0.144
6	そば・うどん	-0.133		0.036	6.8	15.6	-0.338
7	未婚率		0.123	0.026	5.8	13.0	0.375
8	脂質		0.121	0.016	5.6	55.4	0.250
9	規則食事	-0.120		0.030	5.6	58.9	-0.549
10	初婚年齢		0.090	0.021	3.1	28.9	0.284
11	出生率	-0.089		0.023	3.1	1.46	-0.384
12	飲酒		0.060	0.019	1.4	20.9	0.390
13	受動喫煙		0.060	0.009	1.4	51.0	0.250

太字：これまで未検証の新規要因

影響要因を含む全説明変数について死亡率との単相関係数と死亡率に対する感度との散布図を図3に示す。この図をみると、全説明変数の中で感度が高いものとして選ばれた影響要因の中には相関係数の絶対値が低いものがある。また逆に、説明変数の中には相関係数が高いにもかかわらず、感度が低く影響要因にならなかったものが多数ある。さらに、感度と相関係数が異符号の変数も多数ある。これらの結果は、感度が死亡率に対する説明変数個々の正味の感度を示すのに対し、死亡率との相関係数には他の説明変数の寄与が含まれているためであるが、先行研究における解析に関して2つの問題点を示唆す

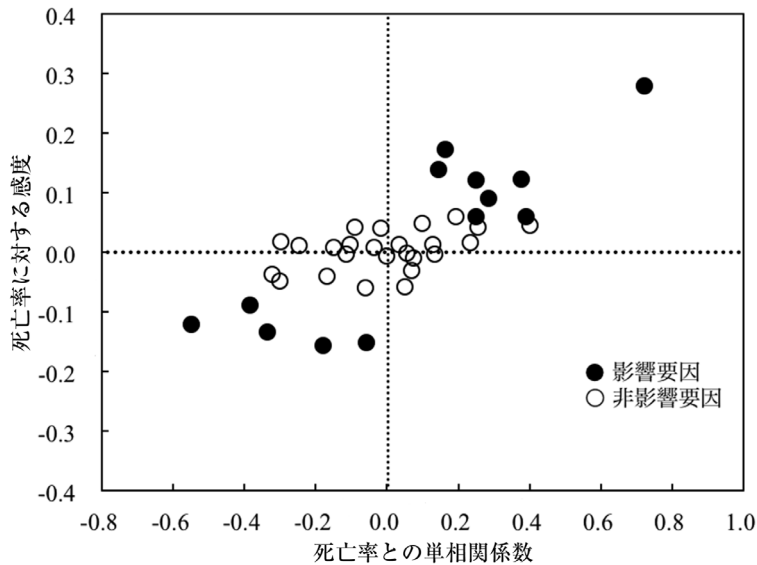


図3 全説明変数についての死亡率との単相関係数vs感度の散布図

る。

第1は死亡率との相関分析に基づく説明変数の影響度の考察である。これまでは都道府県の死亡率との相関係数に基づいて各説明変数の影響度を分析した研究が多い。しかし、本研究の結果はこのような相関分析で得られた先行研究の結果には信頼性の点で疑問があることを示唆する。第2は説明変数の選定であり、死亡率と相関の高い指標を説明変数に選び解析した先行研究が多い。しかし、このような方法では選定された説明変数以外に死亡率に大きな影響を与える変数を見落とす可能性がある。たとえば、影響要因4位となった運動は相関係数が-0.059と非常に低いため、説明変数として見落とす可能性が高い。

以上の結果を総括すると、先行研究では説明変数の選定や死亡率との影響度に関する考察において信頼性に疑問がある場合が考えられる。これに対し、本研究では死亡率に対する正味の感度に基づいて広範な分野の多数の説明変数の中から探索した影響要因が死亡率を統計的に有意な精度で再現する結果が得られたことから、本研究の解析法の有効性が実証されたと考えられる。本研究のように多分野の多数の説明変数を用いて解析し、影響要因を探索した研究は見当たらない。

4. 考察

次に、本研究で得られた影響要因について既往の評価研究の結果(表1)と比較すると、要因によって乳がんに対する影響度に違いがあることが分かる。

① 喫煙

まず、本研究では喫煙が影響要因1位となり、乳がん死亡率に30%という高率で寄与することを示す結果が得られたが、既往の評価研究の結果とは大きく異なっている。すなわち、煙草の煙には60種以上の発がん物質が含まれていることからIARCは限定的証拠ありと評価している。また、日本乳癌学会はリスクがほぼ確実、国立がん研究センターはリスクの可能性ありと評価している(表1のWCRF/AICRと癌研究会の空欄は評価していないことを示す)。これらの評価結果の違いはコホート研究や症例対照研究の手法を用いて喫煙と乳がんの関連を調べた疫学論文の結果が一致していないためである。これら既往評価研究は全体として乳がんに対する喫煙のリスクは小さいと評価しており、本研究の喫煙の影響がきわめて高いという結果とは大きな違いがある。

なお、受動喫煙については、本研究では男性の喫煙率を近似値と仮定して用いたが、影響要因13位、寄与率1%であり、乳がん死亡率への影響は非常に小さいことを示す結果となった。既往の受動喫煙についての評価研究は日本乳癌学会、国立がん研究センターともに女性乳がんに対してリスクの可能性ありと評価している。

② 肥満

肥満が乳がんに与える影響については、既往の評価研究は各種要因の中で最も高い危険度があると評価している。すなわち、WCRF/AICRは肥満が閉経後の女性の乳がんリスクを確実に増加させると評価している(ただし、閉経前ではほぼ確実に「減少させる」としている評価は我々日本人には理解しがたい)。国立がん研究センターは日本人の疫学研

究の結果から、閉経前の危険度は可能性あり、閉経後の危険度はほぼ確実であると評価し、日本乳癌学会も同程度に評価している。このように既往の評価研究が全体的に肥満のリスクを非常に高いと評価しているのに対し、本研究の結果では肥満は影響要因2位になったものの、乳がん死亡率への寄与は1位の喫煙の30%よりはるかに低い10%に過ぎず、既往評価研究の評価とはかなりの違いがある。

③ 乳がん検診

わが国では2000年からマンモグラフィによる乳がん検診が実施されているが、その罹患・死亡への効果についての疫学研究は少ないため、既往の評価研究の評価はない。がん検診の受診率の全国平均35%は欧米の70~80%より大幅に低く、また都道府県格差が26~49%と著しい。都道府県別の乳がん死亡率について多数の説明変数を用いた重回帰分析により、乳がん死亡率に対するがん検診の影響度を示したのは本研究が初めてである。本研究で乳がん検診が死亡率にある程度の効果があることが実証されたことから、受診率の向上が望まれる。

④ 運動

運動が乳がんに与える効果を調べたコホート研究によれば、閉経後の女性では適度な運動が乳がんのリスクを下げることから、WCRF/AICRと日本乳癌学会は閉経後の乳がんにはほぼ確実と評価し、国立がん研究センターは可能性ありと評価している。しかし、閉経前女性については多くの疫学研究の結果が一定していないことから、WCRF/AICRと国立がん研究センターは可能性ありと評価し、日本乳癌学会は証拠不十分と評価している。これらの既往評価研究の結果に対し、本研究では運動の乳がん死亡率に対する影響度は4位の9%の寄与率であり、既往評価研究の結果にはほぼ対応している。

⑤ 牛乳・乳製品、脂質

牛乳には蛋白質、脂質、炭水化物、ビタミン、ミネラルなど、乳がんのリスクを増加させる成分と減少させる成分が含まれるため、牛乳・乳製品と乳がんの関係については多くの疫学研究が行われてきたが、それらの結果は一致していない。そのため、WCRF/AICRでは閉経前は予防効果の可能性あり、閉経後は証拠不十分と評価し、日本乳癌学会は予防効果の可能性あり、国立がん研究センターは証拠不十分と評価している。これらの既往評価研究が全体としてはリスク減少と評価しているのに対し、本研究では牛乳・乳製品は乳がんの死亡率に対する感度が正で、乳がんのリスク増加を示す結果になり、既往評価研究とは効果の方向が正反対である。

一方、脂質については動物実験や多くの疫学研究が行われてきたが、それらの結果は一致していない。そのため、WCRF/AICRでは閉経前は証拠不十分、閉経後はリスクの可能性ありと評価し、日本乳癌学会と国立がん研究センターは証拠不十分と評価している。これに対し、本研究では脂質摂取は乳がんのリスク増加を示し、寄与率が非常に低いので、既往評価研究とほぼ一致している。

⑥ そば・うどん

穀類の摂取が乳がんに与える影響を調べた疫学研究が非常に少ないため、既往の評価研究は証拠不十分と評価している。これに対し、本研究の結果はそば・うどんの摂取の感度が負で、乳がん予防に効果があること示しているが、このような結果は本研究が初めてで

ある。特にそばは食物繊維、およびルチンというポリフェノールを多量に含んでおり、食物繊維ががん予防に効果があるという報告が多いことから、そばの摂取と乳がんの関係を調べる疫学研究が待たれる。

⑦ 未婚率，初婚年齢，出生率

乳がんの発症には女性ホルモン（エストロゲン）が関連しており，体内のエストロゲン濃度が高い期間が長いほど発症リスクが増加する。したがって，初経から閉経までの期間が長い，出産歴がない，または数が少ない，初産年齢が遅い，授乳歴がない，または期間が短い等の要因が乳がんのリスクとなる。既往の評価研究では乳癌研究会と日本乳癌学会がこれらの要因の評価を行っているが，本研究では未婚率，初婚年齢，出生率のいずれも影響要因となり，既往評価研究の結果とほぼ一致している。

⑧ 規則食事，バランス食事，睡眠，ストレス，健康無関心

乳がんの罹患・死亡の低減には，これまでの疫学研究で検証されてきた生活習慣や摂取食品，出産関連要因だけでなく，普段の生活で健康に関心・意識をもち，積極的にがん検診を受診することが肝要であると考えられる。しかし，このような健康に関する心理的要因が乳がんに与える影響を調べた研究は少なく，既往評価研究の中では日本乳癌学会が心理ストレスについて十分な証拠なしと評価しているのみである。本研究では，規則食事（規則正しく食事をとっている女性の比率），バランス食事（バランスの取れた食事をしている比率），睡眠（睡眠を十分にとっている比率），ストレス（ストレスを抱えている比率），および健康無関心（健康のために特に何もしていない比率）の新規5要因を説明変数に採用して解析した。その結果，規則食事が影響要因の9位に入り，6%弱ではあるが，乳がん死亡率に寄与していることを初めて明らかにした。これらの心理的要因の影響についても疫学研究が待たれる。

⑨ 飲酒

アルコール飲料に含まれるエタノールとその代謝物であるアセトアルデヒドには発がん性があることが実験的に示され，IARCは危険度確実と評価し，WCRF/AICRは閉経前ではほぼ確実，閉経後は確実と評価している。しかし，日本乳癌学会は，欧米人と比べて日本人女性は飲量が少ないことからWCRF/AICRの評価を1ランク下げ，閉経前は可能性あり，閉経後はほぼ確実と評価している。一方，国立がん研究センターは日本人についてのデータが少ないことから証拠不十分と評価している。これらの既往研究の評価に対し，本研究では飲酒は影響要因に入ったものの，順位は12位，寄与率はわずか1%であり，既往評価研究の評価よりは危険度が低目の結果となった。

⑩ 大豆製品

乳がん罹患率が欧米人で高く，アジア人で低いことの理由として大豆製品がアジアで多く摂取されていることが挙げられている。欧米では大豆製品の摂取に関するデータが少なくWCRF/AICRは証拠不十分と評価しているが，日本乳癌学会と国立がん研究センターは日本人を対象にした研究をもとに予防効果の可能性ありと評価している。本研究では大豆製品摂取量を説明変数に採用して解析したが，影響要因に入らず，日本人対象の評価研究と研究と異なる結果となった。

以上の本研究と既往の評価研究との比較を総括すると、全般的には本研究の結果は既往の評価にほぼ対応していると考えられる。しかし、本研究の結果の内、喫煙の影響度が非常に高い、肥満や飲酒の影響度がかなり低い、牛乳・乳製品の影響が逆方向である、などの点で既往評価研究と異なった結果が得られた。これらの違いの原因は不明であるが、欧米人と日本人では喫煙・飲酒・食品摂取の量が異なることや、食品摂取量の正確な把握が困難なことも関係していると考えられ、今後、より詳細なデータを活用した総合的な解析が必要である。

最後に、本研究の結果に基づいて地域の乳がん死亡率低減対策に資する情報提供の可能性を考察してみる。考察の対象地域として、国内で乳がん死亡率が長年にわたって最大の青森県と、最低の鳥根県を取り上げると、両県での影響要因13種の国内順位は表4のようになる。ここで各要因の危険度順位は、感度の符号が正の危険要因については指標値の降順、符号が負の予防要因については昇順の順位である。この表を見ると、鳥根県は全体的に危険度順位が低い要因が多く、死亡率が全国最低であることがわかる。これに対し、青森県は危険度順位の高い要因が多く、特に影響度の高い喫煙と肥満の危険度順位の高さが青森県の死亡率の国内最高に大きく貢献していることが明らかである。

表4 影響要因の最小県と最大県、および青森県と鳥根県の危険度順位

影響要因	影響方向	最小県		最大県		青森県		鳥根県	
		数値	該当県	数値	該当県	数値	順位	数値	順位
喫煙	↑	6.1	徳島	17.8	北海道	14.3	2	6.3	46
肥満	↑	21.9	京都	23.7	沖縄	22.9	2	22.2	34
乳がん検診	↓	26.4	山口	48.6	山梨	31.7	13	30.7	7
運動	↓	25.0	青森	38.6	東京	25.0	1	31.0	14
牛乳・乳製品	↑	84	佐賀	137	東京	115	24	129	4
そば・うどん	↓	8.2	沖縄	30.0	香川	14.0	14	12.9	10
未婚率	↑	8.7	福井	19.2	東京	13.9	14	11.1	37
脂質	↑	49	高知	60	東京	55	23	53	42
規則食事	↓	48.9	沖縄	64.3	長野	54.1	2	63.0	46
初婚年齢	↑	28.1	福島	30.3	東京	28.7	27	28.6	34
出生率	↓	1.09	東京	1.90	沖縄	1.36	10	1.68	46
飲酒	↑	16.4	三重	29.0	東京	21.4	15	18.1	40
受動喫煙	↑	44.2	沖縄	58.9	青森	58.9	1	50.8	24

↑：危険要因， ↓：予防要因， 順位：危険度順位

青森県では現在、健康づくりの推進や自殺対策などの短命県返上への模索を実施しているが、本研究の影響要因の結果を基に青森県の乳がん死亡率改善策を考えてみると、まず、喫煙対策がある。死亡率に対する喫煙の影響度は1位と高いが、青森県の女性の喫煙率は全国で2番目に高い14.3%であり、これは最低の徳島県の2倍以上である。また、影響度2位の肥満と4位の運動の改善も有効であろう。運動の順位（国内1位）を改善することで肥満の順位（国内2位）も改善される可能性が高い。

次に産業転換策を考えてみる。青森県といえば津軽のりんごと大間のマグロが名高い農

業・漁業県であり、第1次産業就業者率は全国1位である。また、同県は降雪日数が2位の雪国県である。これらの社会的・自然的背景要因が喫煙率2位、県民所得、大卒率、屋外活動率、インターネット利用率等の最下位をもたらし、さらにこれらの要因が健康意識の低さ、無関心に繋がり、最終的に乳がんの死亡率の高さを招いていると推測される。

そこで、青森県の根本的な死亡率低減対策には健康意識の向上が不可欠であるが、そのための一つの方策として第1次産業からの産業構造の転換を提案する。同県が雪国県である点を考慮すると、情報通信業への転換が有効と思われる。その見本になるのは自然条件が似ている北欧のフィンランドとエストニアであり、両国の新興IT企業は世界に進出している。青森県内の大学や高専の情報関係学部と奨学金制度の拡充強化やIT企業の誘致に成功すれば、女性の教育水準の改善が図られ、これらを通じて健康意識の向上、ひいては乳がん死亡率の低下が期待できよう。

5. 結論

本研究では、地域の乳がん対策に資する情報提供の可能性を探る観点から、都道府県別の乳がん死亡率を目的変数とし、40種の指標を潜在的な説明変数の候補として非線形重回帰分析により解析する実証研究を試みた。その結果、乳がん死亡率を統計的に有意な精度で再現する13種の影響要因が得られ、その中では喫煙や肥満が乳がん死亡率に高い影響を与えることを見出した。また、飲酒や牛乳・乳製品および大豆摂取量等の一部の影響要因については、これまでの疫学研究の結果とは異なる結果が得られた。さらに、乳がん死亡率が国内で突出して高い青森県の死亡率低減策を検討し、いくつかの提言を試みた。

しかし、本研究の結果には多くの課題が残されている。第一は、本研究の解析手法は生態学的研究であるため、本研究で得られた結果は乳がん死亡率と要因との関連性を示すものであって直接的な因果関係ではない。乳がん罹患の一次予防に用いるデータは因果関係を示すものでなければならず、そのためには動物実験や疫学的研究が必須である。しかし、二次予防のための乳がん検診の受診対象者選別には関連性を示すもので十分である。そこで、個人単位のマイクロデータ等の各種データを利用した総合的な解析を行ない、本研究の結果の精度を高めることで二次予防への応用が可能であると考えられる。

第二の課題は、本研究では40種の説明変数のみを潜在的な要因として取り上げたが、WHOの報告書等によれば、人の死亡には本研究で取り上げた指標以外に、職業、医療、住居、環境、人間関係、文化活動等、多くの経済・社会的背景要因が挙げられ、しかもこれらの要因が個人から集団まで複雑な階層構造を形成して相互に関連して死亡に影響するとされている。例えば、欧米では専門技術的職業等の社会階層の上位の女性の方が乳がんにかかりやすいことが報告されているが、社会階層は妊娠、出産、生活習慣等の要因を介して乳がんに関係している可能性が高い。さらに、妊娠、出産等の要因は生活習慣と密接な関連があり、これらを介して乳がんに関係している可能性がある。このような様々な階層構造を考慮した乳がんの要因分析は今後の検討課題である。

参考文献

- 1) World Cancer Research Fund/American Institute for Cancer Research, "Diet, Nutrition, Physical Activity and Cancer: A Global Perspective. Continuous Update Project Expert Report 2018," <https://www.wcrf.org/sites/default/files/Summary-third-expert-report.pdf>.
- 2) International Agency for Research on Cancer, "IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 83, Tobacco Smoke and Involuntary Smoking," <https://monographs.iarc.fr/wp-content/uploads/2018/06/mono83.pdf>; Volume 96, Alcohol Consumption and Ethyl Carbamate," <https://monographs.iarc.fr/wp-content/uploads/2018/06/mono96.pdf>.
- 3) 富永祐民, 吉田譲, 高谷治, 石田常博, 七沢武, 坂元吾偉, 平山雄, 廣畑富雄, 村田紀, 渡辺昌, 黒石哲生, 加藤育子「乳癌のhigh risk groupとは」乳癌の臨床, **3**, 251-257(1988) .
- 4) 日本乳癌学会編「乳癌診療ガイドライン2 疫学・診断編 2018年版」金原出版(2018) .
- 5) 国立がん研究センター「科学的根拠に基づく発がん性・がん予防効果の評価とがん予防ガイドライン提言に関する研究」班「エビデンスの評価」https://epi.ncc.go.jp/cgi-bin/cms/public/index.cgi/nccpei/can_prev/outcome/index.
- 6) 大北剛 (訳)「サポートベクターマシン入門」共立出版 (2005) .
- 7) 小野田崇「サポートベクターマシン」オーム社 (2007) .
- 8) 阿部重夫「パターン認識のためのサポートベクトルマシン入門」森北出版 (2011) .
- 9) Chih-Chung Chang and Chih-Jen Lin, "LIBSVM-A Library for Support Vector Machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- 10) Kazutoshi Tanabe, Takio Kurita, Kenji Nishida, Bono Lucic, Dragan Amic, and Takahiro Suzuki, "Improvement of Carcinogenicity Prediction Performances Based on Sensitivity Analysis in Variable Selection of SVM Models," *SAR and QSAR in Environmental Research*, **24**(7-9)565-80 (2013) .
- 11) 田辺和俊, 鈴木孝弘「サポートベクターマシンを用いた世界各国の平均寿命の決定要因の実証分析」厚生指標, **61**(13)23-30 (2014) .
- 12) 田辺和俊, 鈴木孝弘, 中川晋一「サポートベクター回帰による都道府県別肺がん死亡率の関連要因に関する検討」保健医療科学, **65**(6)598-610 (2016) .
- 13) 田辺和俊, 鈴木孝弘「平均寿命および健康寿命の都道府県格差の解析 - 非線形回帰分析による決定要因の探索 -」季刊社会保障研究, **51**(2)198-210 (2016) .
- 14) 田辺和俊, 鈴木孝弘「出生率の都道府県格差の分析」厚生指標, **63**(5)13-21 (2016) .
- 15) 田辺和俊, 鈴木孝弘「都道府県別全がん死亡率に及ぼす生活習慣要因の影響度分析 - 自治体のがん対策の視点から -」厚生指標, **65**(11)15-21 (2018) .