

一般口演

一般口演3

医療データ分析1（臨床データ解析）

2018年11月23日(金) 10:15～11:45 G会場 (5F 504+505)

[2-G-1-2] DPCデータを用いた決定木分析による乳がん和生活習慣病との関連について

○坂本 千枝子¹, 外山 比南子², 斎藤 恵一¹ (1.国際医療福祉大学, 2.医療データサイエンス研究所)

【目的】がんは現在我が国の死因第1位であり、喫煙や生活習慣病がそのリスク要因になっているとの報告も多い。そこで、主傷病や併存症の他、種々の情報を利用できる DPCデータを利用して2013年以来5大がん（肺、胃、肝、大腸、乳房）と喫煙や生活習慣病との関連を調べてきた。今回は乳がんを対象に同様の検討を試みる。【方法】2010-12年の4病院の DPCデータ18,448件の内訳は、複数回入院を1患者としてみると全体で14,780件になった。調査項目に欠落があるデータを除外すると、がん疾患は2,863件で乳がんは121件だった。がん疾患を乳がんとそれ以外のがんの2群に分類した。SPSSの決定木手法を用いて従属変数に乳がんの有無、独立変数に高血圧(HT)、高脂血症(HL)、2型糖尿病(DM)、肥満指数(BMI)、入院時年齢（歳）、喫煙の有無（喫煙）を置き、分岐がなくなるまで最も関連の強い第1分岐項目を独立変数から除いていく方法で分析した。【結果】第1分岐は歳で53以下、53超67以下、67超75以下、75超の4群に分かれ、乳がんの割合は歳が若いほど高かった。75超以外は第2分岐まであり、すべて喫煙だった。次に歳を除くと第1分岐は喫煙で喫煙有はHLで分岐し、HL有は無より乳がんの割合が約3倍だった。さらに喫煙を除くとDMで分岐し、DM無は有より乳がんの割合が約2倍だった。【考察】第1分岐の年齢群は乳がんの好発年齢が50代前半との報告を裏付ける。また、国際評価（IARC）では喫煙は乳がんのリスクとされていないが、HL有だと乳がんの割合が高くなったことは注目すべき点と考える。DMではリスク要因とされる報告と逆の結果となったが、治療薬が発がんリスクを抑えるとの報告もあり、調査が必要と考える。乳がんでは生活習慣病の関連がHTよりもHLやDMがより大きいことが示唆された。【結語】DPCデータを利用して決定木手法を用いた分析の結果、乳がん患者の年齢層や喫煙有とHL有の関連、DMとの関連が示唆された。

DPCデータを用いた決定木分析による乳がん和生活習慣病との関連について

坂本 千枝子^{*1}、外山 比南子^{*2}、
齋藤 恵一^{*1}

*1 国際医療福祉大学大学院、*2 医療データサイエンス研究所、

The influence of lifestyle-related disease on breast cancer by the decision tree analysis using DPC dataChieko Sakamoto^{*1}, Toyama Hinako^{*2}, Keiichi Saitou^{*1}

*1 International University of Health and Welfare Graduate School, *2 Institute of Medical data Sciences

Abstract : We investigated the relations among age, smoking information, and lifestyle-related disease (hypertension :HT, type 2 diabetes ;DM, hyperlipidemia : HL, body mass index: BMI) in cases with breast cancer, using Diagnosis Procedure Combination survey data (DPC data). The risk factors to affect breast cancer was investigated using a method of decision tree. As a result, age was extracted as the most significant factor relating to breast cancer. The rate of breast cancer patients with the youngest age group was 10 times higher than the rate of the oldest age group. The second factors was smoking. The rate of no-smoking groups are higher than the rate of smoking group. After the age was excluded from an independent variable, a smoking factor was extracted as the most significant factor. The ratio of the breast cancer in no-smoking group was more than that of smoking group in which the following factor was HL. The rate of breast cancer patients with HL was 3times higher than the rate of patients without HL. After the smoking factor was excluded from an independent variable, the rate of breast cancer patients without DM was 2times higher than the rate of patients with DM. patients have most HT. It was suggested that HL and DM are more lifestyle-related disease strongly affect on the breast cancer than HT which is most associate disease.

Keywords: breast cancer, lifestyle-related disease, Decision tree, DPC data

1. 緒論

がんは1981年以来わが国の死因第1位であり、生活習慣病の併存が、がん発症のリスク要因になっているとの報告も多い。DPCデータには患者データが集約されており、様式1だけでも主傷病の経過や予後に影響する併存症はじめ様々な情報を利用することができる。そこで、協力病院の2010-2012年のDPCデータを利用してこれまでがんの発症や治療、予後等臨床に役立つ項目として併存症やリスクファクタを検討してきた。院内がん登録の5大がん(肺、胃、肝、大腸、乳房)についてそれぞれ上位4位までの併存症を調べた結果、生活習慣病の高血圧(HT)と2型糖尿病(DM)は5大がんすべてに、高脂血症(HL)は肺がん、胃がん、乳がんに併存することがわかった¹⁾。そこで、2013年以来、5大がん和生活習慣病やその他のがんリスクとされる項目との関連を調べてきた。

2. 目的

DPCデータを利用し、乳がん和生活習慣病やその他のがんリスクとされる項目がどのように関連しているか検討する。

3. 方法**3.1 調査項目**

我が国の5大がんのうち、近年罹患数、死亡数ともに上昇している乳がんを対象疾患とした。乳がんに関連する要因を検討するために、生活習慣病として対象データの5大がんそれぞれの上位4位までに入った併存症のHT、HL、DMを選択した。この他に、代表的な乳がんのリスクファクタを「乳がんガイドライン2 疫学・診断編2015年版」²⁾を参考にして、入院時年齢(年齢)、肥満指数(BMI)、喫煙の有無(喫煙)を選択した。

3.2 対象データの概要

2010-2012年のDPCデータのうち20歳以上の患者データ18,448件を利用した。入院を繰り返している複数回入院も1患者としてみると全体で14,780件になった。BMIと喫煙が不明の症例を除外した結果、全体で13,252件となった。このうちがん疾患ありは2,863件となり、その内訳は乳がん121件、胃がん200件、肝がん217件、大腸がん284件、肺がん320件、その他のがんは1,721件だった。この内訳を表1に示す。

表1 対象データの内訳

疾患内訳	件数	全患者に 占める割合(%)	全がん患者に 占める合(%)
全患者数	13,252	—	—
がん疾患なし	10,389	78.4	—
がん疾患あり	2,863	21.6	—
乳がん	121	0.9	4.2
胃がん	200	1.5	7.0
肝がん	217	1.6	7.6
大腸がん	284	2.1	9.9
肺がん	320	2.4	11.2
その他	1,721	13.0	60.1

3.3 分析方法

がん疾患を乳がんとそれ以外のがんの2群に分類した。分析には統計ソフトIBM SPSS Statistics23の決定木手法(CHAIID)を用い、従属変数に乳がんの有無、独立変数にH

T、HL、DMと年齢、喫煙、BMIを設定し、分岐がなくなるまで最も関連の強い第1分岐項目を独立変数から除いていく方法で分析した。

3.4 倫理的配慮

使用したDPC調査データは、病院名もデータ識別番号も連結不可能匿名化されており、個人情報を含んでいないため、国際医療福祉大学研究倫理審査において「非該当」とされた(承認番号:14-Ig-146)。

4. 結果

4.1 生活習慣病、年齢、BMI、喫煙の有無の内訳

対象症例の5大がんの併存症である生活習慣病と年齢、BMIについては、第36回医療情報学連合大会で発表した結果³⁾を引用する。乳がんはHT併存有の割合が18.2%、DM併存有の割合が7.4%、HL併存有の割合が13.2%だった。他の4部位に比べてHLを併存する割合が最も高かったが、逆にHTとDMを併存する割合は最も低かった。年齢の中央値は乳がんが61歳と最も低く、肝がんが73歳で最も高かった。乳がんと肝がんはそれぞれ、その他のがんと間にMann-WhitneyのU検定($p < 0.05$)で有意差がみられた。BMIの中央値は乳がんが22.3で、肝がんが22.8で最も大きく、肺がんが21.7で最も小さかった。肝がんと肺がんのBMIの中央値はそれぞれ、その他のがんと間にMann-WhitneyのU検定($p < 0.05$)で有意差がみられた。しかし、乳がんを含めて中央値が肥満を示す25以上の部位はなかった。喫煙状況について表2に示す。肺がんのみ喫煙有の割合が59.4%で最も高かった。胃がん、肝がん、大腸がんの喫煙有が40%台だったのに比べて、乳がんは15.7%と最も低かった。

表2 5大がんの喫煙状況

	喫煙有		喫煙無	
	人	%	人	%
乳がん	19	15.7	102	84.3
胃がん	88	44.0	112	56.0
肝がん	106	44.8	111	51.2
大腸がん	116	40.8	168	59.2
肺がん	190	59.4	130	40.6

4.2 決定木手法による分析

従属変数に乳がんの有無を設定し、HT、HL、DM、年齢、BMI、喫煙を独立変数に設定したツリー全体図を図1に示す。第1分岐は年齢で53歳以下、53歳超67歳以下、67歳超75歳以下、75歳超の4群に分かれた。乳がんの割合は歳が若い群ほど高く、53歳以下は75歳超の約10倍だった(図1:ノード1~4)。75歳超以外は第2分岐まであり、すべて喫煙の有無だった。乳がんの割合は、いずれも喫煙無で年齢の若い群ほど高かった。(図1:ノード5~10)。乳がんの割合は、53歳以下で喫煙無は67歳超75歳以下で喫煙無の約4倍だった(図1:ノード5、ノード9)。

次に、初めに設定した独立変数の項目から、乳がんと最も関連が強いと考えられる年齢を除いた結果を図2に示す。第1分岐は喫煙の有無で、喫煙無が有より乳がんの割合が約4倍だった。しかし、喫煙有はさらにHL併存の有無で分岐し、

HL有は無より乳がんの割合が約3倍だった(図2:ノード0~4)。

同様に、独立変数から喫煙の有無を除いた結果を図3に示す。分岐はDM併存の有無の1回だった。DM無は有より乳がんの割合が約2倍だった(図3:ノード0~2)。

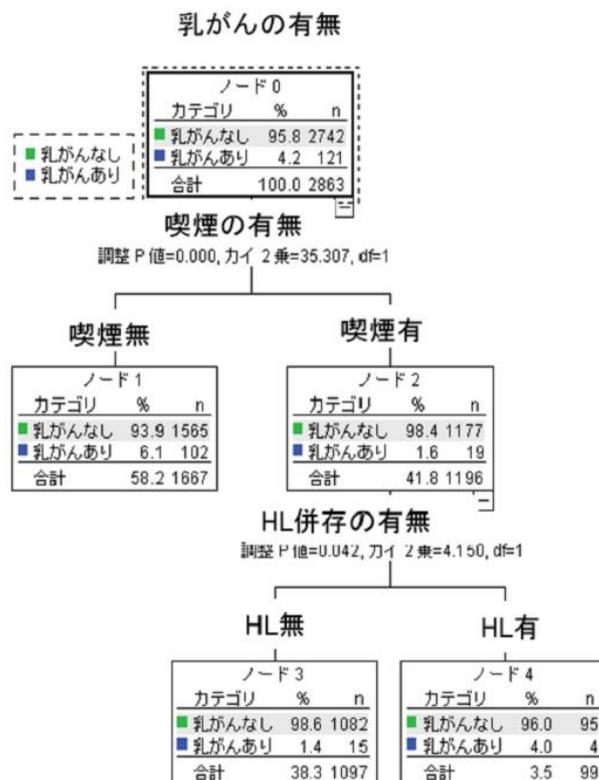


図2 乳がんの有無 ⇒ 喫煙の有無
⇒ 喫煙有 ⇒ HL 併存の有無 による分岐

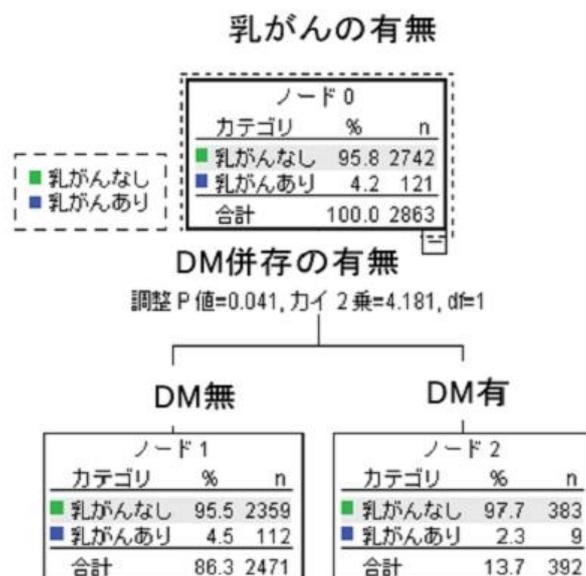


図3 乳がんの有無 ⇒ DM 併存の有無 による分岐

5. 考察

地域がん登録全国推計によるがん罹患データの全国年齢階級別推定罹患数、部位、性、診断年別(1975年～2013年)⁴⁾によると、2010年～2012年の乳がんの罹患数は60歳～64歳が最も多く、次いで45歳～49歳、65歳～69歳となっている。第1分岐の年齢群と同様の分布であることから、年齢構成において対象データを乳がんの標準的な集団とみなすことができると考える。そして、年齢群が若いほど乳がんの割合が高いことは、乳がんの罹患率が30歳台後半から増加し始め、40歳台後半から50歳台前半でピークを迎え、その後は次第に減少する⁵⁾を裏付けている。また、国際評価(IARC. 米国がん研究協会 第2版, 2007年)では喫煙は乳がんのリスクとされていなかったが、日本人のエビデンスに基づく評価の現状についての国立がん研究センターの研究班による「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」⁶⁾では、「可能性あり」とされている。対象データでは第2分岐がどの年齢群も喫煙の有無で、そのすべてで喫煙無群の方が乳がんの割合が高かった。喫煙有群が全体の15.7%だったことが影響していると考えられる。喫煙の発がんリスクについては受動喫煙も見逃すことはできないこともあり、この分岐に関してはこの対象データだけで判断することの限界といえる。しかし、最も関連の強いと考えられる年齢を独立変数から除いたところ、第1分岐の喫煙の有無では、喫煙無で乳がんの割合が高いことは同様だったが、喫煙有はさらにHL併存の有無で分岐した。そして、乳がんの割合はHL有が無の約3倍になった。対象データでは、5大がんのうち乳がんがHLを最も高い割合で併存していた。メバロチンやリビドールなどのスタチンはHLの第一選択の治療薬である。様々な癌腫でスタチン服用と癌発症リスクの関連性が検討されている⁷⁾が、「乳がんガイドライン2 疫学・診断編2015年版」²⁾のCQ26では、エビデンスグレードが**Limited-no conclusion**(証拠不十分)で、スタチンの服用が乳癌発症リスクを減少させるかどうかは結論付けられないとされている。治療薬の影響の有無に関わらず、HLの併存は乳がんとの関連を示唆すると考える。次に喫煙の有無を独立変数から除いたところ、DM併存の有無で分岐し、乳がんの割合はDM無で有の約2倍になった。「乳がんガイドライン2 疫学・診断編2015年版」²⁾のCQ25では、エビデンスグレードが**Probable**(ほぼ確実)で、DMの既往が乳癌発症リスクを増加させることはほぼ確実であるとされている。一方、インスリン抵抗性を改善させるピグリアナドイド薬であるメホルミンについては、メホルミン使用者は非使用者に比較して癌罹患リスクが低いことも報告されている⁸⁾。今後は、併存症の治療薬についても注目していく必要があると考える。HTは、乳がんの併存症の中で最も高い割合だったが、乳がんとの関連は見られなかったことから、乳がんと生活習慣病の関連ではHTよりもHLやDMがより大きいことが示唆された。肥満は乳がんのリスク要因とされている²⁾が、BMIは中央値が22.3だったことで乳がんとの関連は見られなかったと考える。地域がん登録全国推計によるがん罹患データの全国年齢階級別推定罹患数、部位、性、診断年別(1975年～2013年)⁴⁾によると、わが国の2013年の乳がん(上皮内癌を含む)の罹患数の年齢分布は年々増えており、1975年を1とすると1985年が1.8、1995年が2.8、2005年が4.6、2013年が7.7となっている。このことから、乳がんのリスク要因についてさらに調べることは重要である。

6. 結論

乳がんと生活習慣病やその他のがんリスクとされる項目がどのように関連しているか検討することを目的とした。DPC データを利用して決定木手法を用い、従属変数に乳がんの有無、独立変数に併存症の HT、HL、DM と年齢、喫煙、BMIを設定し分析した。第1分岐は年齢で乳がんとも最も強い関連が示唆された。第2分岐は喫煙の有無で、すべての喫煙無群で乳がんの割合が高かった。独立変数から年齢を除くと、喫煙の有無で分岐し、喫煙無の方が乳がんの割合が高かった。喫煙有はさらに HL 併存の有無で分岐し、HL 有が乳がんの割合が高かった。さらに喫煙の有無を除くと DM 併存の有無で分岐し、DM 無の方が乳がんの割合が高かった。乳がんの併存症である生活習慣病では、最も割合の高い HT よりも DM と HL で関連が強いことが示唆された。乳癌診療ガイドラインや文献の乳がんのリスク因子と違った結果もあったが、それぞれの疾患の治療薬の影響も考慮する必要がある。

7. 謝辞

本研究は MEXT 科研費 JP16K00486 の助成を受けたものです。

8. 参考文献

- 1) 坂本千枝子, 外山比南子, 斎藤恵一. DPC データを用いた決定木分析によるがん疾患の要因抽出—肺がん要因の検討—. バイオメディカル・ファジィ・システム学会誌, 2014;16(1):1-6
- 2) 日本乳癌学会編: 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 2015 版, 金原書店, 東京, 2015
- 3) 坂本千枝子, 外山比南子, 斎藤恵一. DPC データを用いた決定木分析による肝臓と生活習慣病の関連について. 医療情報学連合大会論文集, 2016;36(1)454-457
- 4) Hori M, Matsuda T, Shibata A, Katanoda K, Sobue T, Nishimoto H, et al. Cancer incidence and incidence rates in Japan in 2009: a study of 32 population-based cancer registries for the Monitoring of Cancer Incidence in Japan (MCIJ) project. Japanese journal of clinical oncology. 2015; 45(9):884-91.
- 5) <https://ganjoho.jp/public/cancer/breast/> (cited 2018-Aug-23).
- 6) http://epi.ncc.go.jp/can_prev/ (cited 2018-Aug-23).
- 7) Ahern TP, Lash TL, Danker P, et al. Statins and breast cancer prognosis: evidence and opportunities. Lancet Oncol, 2014; 15:e461-468.
- 8) Noto H, Goto A, Tsujimoto T, et al. Cancer risk in diabetic patients treated with metformin: a systematic review and meta-analysis, 2012; PLoS One 7: e33411

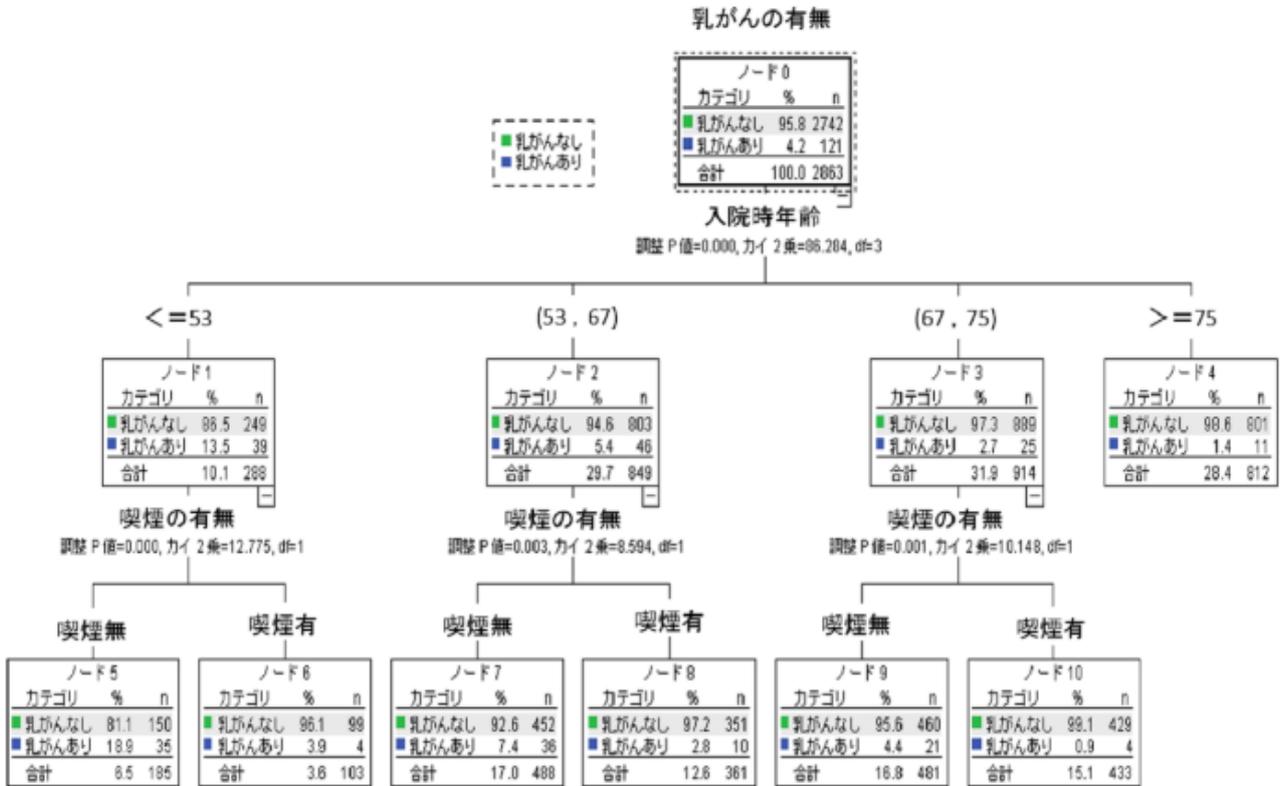


図1 ツリー全体図

乳がんの有無 ⇒ 入院時年齢 ⇒ 喫煙の有無 による分岐