

健診データとがん登録データのレコード・リンケージによる 後ろ向きコホート研究への活用

Retrospective cohort study by the record linkage of health examination and population-based cancer registry

岡本 幹三* 岸本 拓治* 尾崎 米厚* 嘉悦 明彦*

1. はじめに

がん登録データの活用には、罹患状況の把握からがん検診の精度評価、さらには生存率解析による治療方法の評価など多岐にわたる。しかし、がん登録資料のみの活用ではがんの罹患リスクの評価、つまり要因分析にはつながらない場合が多い。こうした状況でがん予防を考える場合、がん発症の予測因子の解明につながる手段として、手近にある市町村における住民健診データの活用は非常に貴重な資料になる。居ながらにして都道府県単位の大規模コホート集団を設定した解析が可能となり、多くのエビデンスを明らかにすることができるからである。

こうした方法が可能になってきたのは、一つには住民健診データの電算化によるデータベース化が可能になってきたことが大きな背景といえる。鳥取県でも、某健診機関で1995年から健診結果の電子化が導入されている。

今回は、鳥取県における健診データとがん登録のレコードリンケージ(記録照合)による後ろ向きコホート研究への活用的一端を紹介しながらどんな利用の仕方ができるのか、どんな問題点や限界があるのかを考えていきたい。まず、図1に健診データの検査項目とその活用例のいくつかを吹出しで示す。

NO.	検査項目	NO.	検査項目
1	市町村コード	22	TCHO
2	検診日	23	TG
3	受診番号	24	HDL
4	個人番号	25	GOT
5	氏名	26	GPT
6	性別	27	GGTP
7	生年月日	28	GUL
8	治療中1	29	A1C
9	治療中2	30	酒
10	治療中3	31	LDL
11	治療中4	32	喫煙状況
12	治療中F	33	開始年齢
13	身長	34	中止年齢
14	体重	35	喫煙本数
15	肥満度	36	喫煙年数
16	標準体重	37	既往歴1
17	最高血圧	38	既往年齢1
18	最低血圧	39	既往歴2
19	食後時間	40	既往年齢2
20	Hb	41	既往歴3
21	Ht	42	既往年齢3

基本健診受診者の
がん罹患特性

治療状況別
がん罹患リスク

肥満とがん罹患

高血圧の
がん罹患リスク

血清脂質と
がん罹患リスク

耐糖能異常と
がん罹患リスク

喫煙状況と
肺がん罹患リスク

「死の四重奏」と
乳がん罹患リスク

図1. 健診データの検査項目と記録照合による活用例

*鳥取大学医学部 社会医学講座 環境予防医学分野
〒683-8503 鳥取県米子市西町 86

2. 基本健康診査受診者のがん罹患特性について

鳥取県で実施している基本健康診査受診者について観察期間を設定し、初回受診時点から観察時点までのがん罹患を鳥取県がん登録資料と照合し同定した。その結果から求められる罹患実測数と一般人口集団のがん罹患率から算出される期待罹患数との比（O/E 比）を求め、性別、部位別の比較をした。その結果、男では前立腺以外、O/E 比は有意に低い値を示し、健診受診者のがん罹患のリスクは低い傾向が認められた。特に呼吸器の肺で継続受診者のO/E 比は0.1以下を示した。女性でも同様の傾向を示し、特に肝臓でO/E 比は0.1以下を示した。healthy screenee effect を考慮に入れなければならないが、健診受診者が健康習慣や健康意識、保健行動が未受診者に比較して良好であることを反映する結果ともいえる。

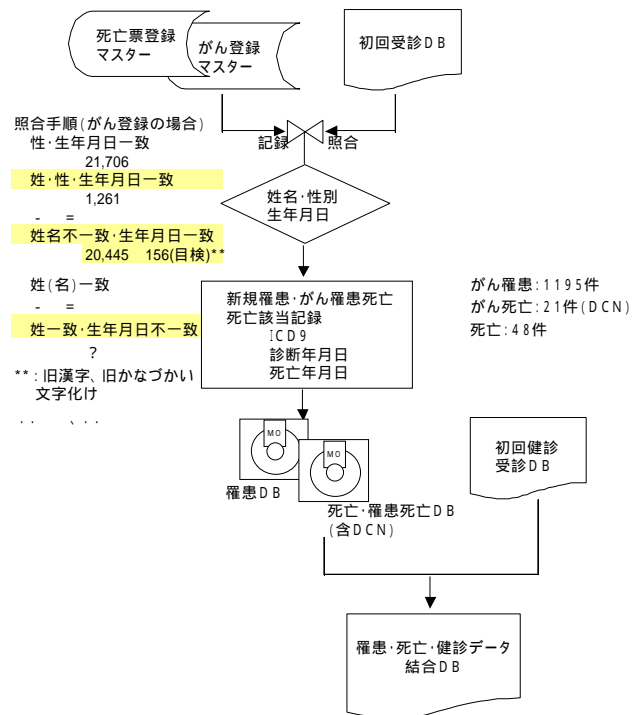


図2. 罹患・死亡記録照合

表1. 継続状況別部位別がん罹患実測数と期待数およびO/E比

	O/E比				O/E比		
	総数	非継続	継続		総数	非継続	継続
全部位	0.49	0.57	0.33	全部位	0.29	0.34	0.20
消化器	0.50	0.59	0.35	消化器	0.30	0.34	0.25
胃	0.66	0.78	0.44	胃	0.34	0.39	0.26
結・直腸	0.39	0.48	0.23	結・直腸	0.33	0.36	0.28
肝臓	0.20	0.22	0.15	肝臓	0.10	0.11	0.08
膵臓	0.61	0.58	0.66	膵臓	0.33	0.43	0.18
呼吸器	0.39	0.53	0.12	呼吸器	0.25	0.28	0.22
肺	0.41	0.61	0.07	肺	0.22	0.21	0.25
前立腺	1.22	1.30	1.08	乳房	0.24	0.29	0.16
泌尿器	0.46	0.39	0.57	女性生殖器	0.26	0.33	0.17
脳、神経	0.55	0.57	0.53	脳、神経	0.24	0.20	0.30
リンパ	0.41	0.54	0.17	甲状腺	0.42	0.60	0.14
造血組織							

3. 基本健康診査受診者の健康診断情報とがん罹患の関係について

これは、健康診断情報を基にしてがん予防に対する新たな知見を得ることを目的に、後ろ向きコホート研究を用いてがん罹患リスクを解析検討するものである。鳥取県においては、これまでに基本健康診査受診者の健康診断情報を活用してがん罹患との関係について検討してきた。以下に5例を紹介する。

(1) 耐糖能異常者とがん罹患リスク

ロジスティック回帰分析により血糖値「正常」を基準にして、「要指導」および「要医療」のがん罹患の相対危険度を算出した。その結果、全部位、結・直腸、膵臓、肺、腎臓で「要医療」や「要指導」の者に有意に高いリスクが見られた。耐糖能異常者は将来がん罹患のハイリスク群と推測され、注意深いフォローアップの必要性が指摘された（表2）。

表2. 耐糖能異常についての部位別がん罹患リスク

	Men			Women		
	obs.	R.R.	95% C.I.	obs.	R.R.	95% C.I.
All Sites	174	2.08	(1.28 - 3.38)	153	2.27	(1.11 - 4.66)
Stomach	59	1.76	(0.75 - 4.15)	45	3.10	(0.94 - 10.10)
Colon	17	2.05	(0.46 - 9.20)	19	2.45	(0.32 - 18.63)
Rectum	15	2.92	(0.63 - 13.54)	14	3.93	(1.08 - 14.33)
Colorectum	32	2.42	(0.83 - 7.05)	33	1.38	(0.19 - 10.20)
Liver	8	2.64	(0.31 - 22.60)	8	5.65	(0.68 - 47.05)
Pancreas	6	3.20	(0.60 - 17.50)	3	5.60	(0.50 - 64.20)
Lung	24	3.92	(1.28 - 12.00)	17	1.93	(0.44 - 8.51)

R.R.: 正常値に対する要医療判定区分の相対危険度(年齢、BMIで調整)

(2) 血清脂質とがん罹患リスク

解析は、Cox 比例ハザードモデルに適用し、初回受診時の総コレステロール、中性脂肪、HDL コレステロールの値をカテゴリー化し、正常値を基準に年齢、BMI、喫煙、飲酒などの交絡因子を投入し調整ハザード比を求めた。その結果、総コレステロールが 149mg/dl 以下でリスクが高く、中性脂肪が 150mg/dl 以上および HDL コレステロールが 61mg/dl 以上でリスクが低下する傾向が観察された（表 3）。この傾向は消化器系がんにおいてのみ観察された。

表 3. 全部位における血清脂質のがん罹患に対する調整ハザード比

		観 察 期 間			
		0～7年		2～7年	
		男	女	男	女
TC	~149	1.21	1.20	1.27	1.15
(mg/dl)	150～219	1.00	1.00	1.00	1.00
	220～	0.80	0.82	0.85	0.90
TG	~149	1.00	1.00	1.00	1.00
(mg/dl)	150～	0.90	0.77	0.89	0.86
HDL-C	~39	1.27	0.82	1.15	0.85
(mg/dl)	40～60	1.00	1.00	1.00	1.00
	61～	0.74	0.86	0.84	0.83
AI	~2.9	1.00	1.00	1.00	1.00
	3.0～	1.07	0.89	1.02	1.03

調整因子：年齢、BMI、飲酒1日3合以上の有無、喫煙習慣（喫煙・非喫煙・禁煙）

(3) 喫煙状況と肺がん罹患リスク

鳥取県がん登録資料を用いて、肺がん検診受診者を喫煙状況別に追跡した後ろ向きコホート研究を行った。その結果、喫煙が肺がんをはじめとするいくつかの部位のがんのリスクファクターになっていること、肺がんに対する喫煙のリスクの大きさは生涯喫煙量が多いほど大きいこと、禁煙者でもそれまでの喫煙量が多いとリスクが大きくなること、禁煙して10年以上経過すると肺がんのリスクは吸わない人と大差がなくなること、等が明らかになった（図 3、4）。

(4) 「死の四重奏」と乳がん罹患リスク

「死の四重奏」（耐糖能異常、高血圧症、高中性脂肪血症、肥満）と乳がん罹患リスクとの

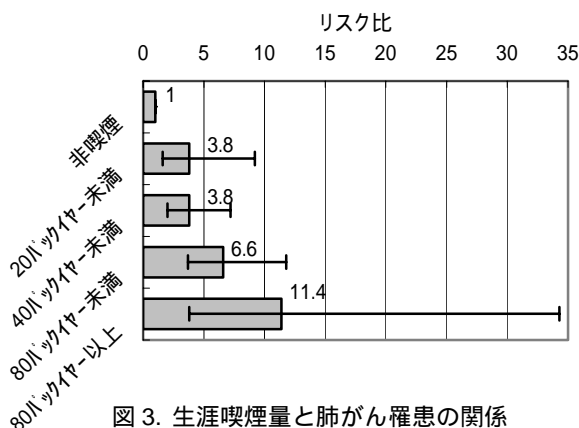


図 3. 生涯喫煙量と肺がん罹患の関係

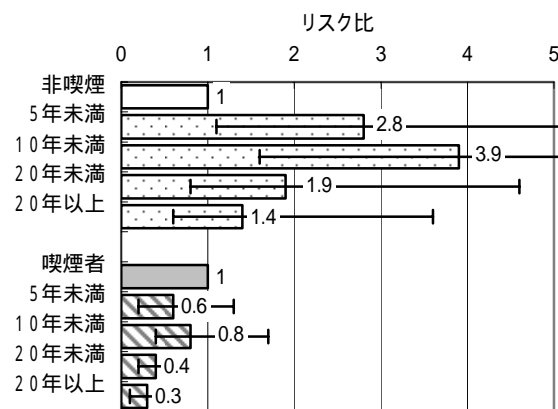


図 4. 禁煙期間と肺がん罹患リスクの関係

関連性を検討するために後ろ向きコホート研究を実施した。その結果、四要因のうち肥満のみが相対危険は 6.37 で統計的に有意であった（図 5）。各個人が有する四要因の個数別の解析では、2 個以上のリスク比は、3.1 で統計的に有意であった。これらの結果は、肥満が、また 2 個以上のリスク要因を有することが乳がん罹患リスクとなり得ることを示唆した（図 6）。

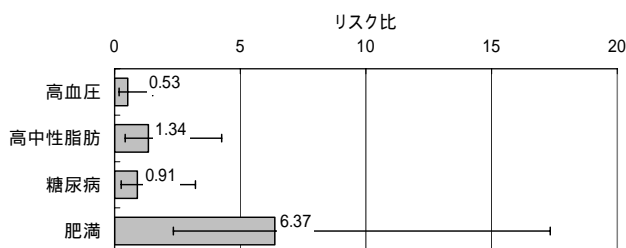


図 5. 乳がん罹患に対する「死の四重奏」要因のリスク比（多変量解析）

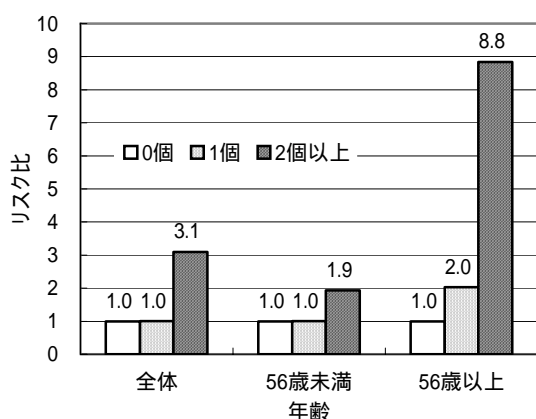


図 6. 乳がん罹患のリスク比 四重奏要因の個数別

(5) 高血圧症とがん罹患リスク

後ろ向きコホート研究を行い、高血圧症とがん罹患リスクとの関連性について検討した。WHO の旧高血圧診断基準に従って、正常群、境界域群、高血圧群に分類した。但し、降圧剤服用者は血圧値によらず高血圧群とした。また、降圧剤の影響を考慮し、降圧剤服用者を治療群として分類し、同様にハザード比を求めた。その結果、いずれも有意な結果は得られず、高血圧とがん罹患の関連性は認められなかった。また、治療群についても有意な結果は得られなかった(図 7)。

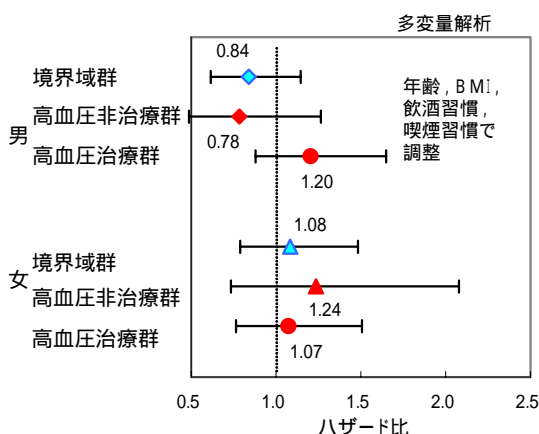


図 7. 高血圧のがん罹患リスク (高血圧治療群と非治療群に分割)

以上、健診データとがん登録の後ろ向きコホート研究への活用例に共通する解析方法の処理手順をまとめてみると、以下の通りとなる。

解析項目：年齢、BMI、血圧、貧血、血清脂質、肝機能、血糖値、心電図判定、眼底判定、喫煙、多量飲酒、治療状況、受診回数、既往歴などで、がん罹患との関連性について検討する。

解析対象：基本健康診査およびがん検診受診者で 40 歳以上のものを対象とする。

観察期間：観察開始は、初回受診年月日(始点)とし、観察終了は、がん罹患日、死亡日もしくは観察終了日(研究班標準集計完了年月日)とする。

記録照合：健診受診者とがん登録患者・死亡者の性、生年月日、姓、名ならびに参考までに住所の一致するものをがん罹患・死亡したものとみなし、その罹患年月日・死亡年月日と罹患部位・死因を転記する。性、生年月日、姓、名不一致についても照合し、見落としをなくす。

解析項目採用条件：健康診断情報はすべて初回受診時の測定記録を採用し、複数がんについては、第 1 がんを採用する。

解析項目除外条件：初回受診時既登録者(既のがんに罹患しているもの)や性・生年月日が不明のものは無効データとして解析から除外する。

治療中の取扱い：治療の影響をみる場合は含めて解析し、治療の影響を除外したリスクを評価する場合は除外する。

解析前処理：連続変数のカテゴリー化ならびに変数カテゴリー値の並び替えをする。基準となるカテゴリー値を最小値にする。例えば、正常を 0、要指導を 1、要医療を 2 とする。カテゴリー化は、カットポイントによって結果が左右されることもあるので分布を見ながら検討する。

解析方法：SPSS によるロジスティック回帰分析、Cox 比例ハザードモデルに適用し健康診断情報に関する罹患リスク(相対危険度、ハザード比)を算出する。

4. 健診データの活用をめぐる問題と限界

既存資料を利用する関係で、後ろ向きコホート研究の短所といえるかもしれないが様々な制約があるのは避けられない。以下にその例をあげて説明する。

(1) 生活習慣情報の欠如

基本健診において、生活習慣情報が記録されていない場合が多々ある。鳥取県もその例にもれずとくに食生活、運動、睡眠、休養に関する情報が欠損しているのが問題である。がん罹患に係わる要因分析が出来ないことと、交絡因子としての罹患リスクの調整が困難となることである。

(2) 既存資料(過去の基本健診データ)の外部提供をめぐる諸問題

個人情報保護の立場から市町村委託の検診機関からの健診データの提供に関する市町村の承諾を得ることは年々難しくなっている。鳥取県においても同様で、個人情報保護法の成立の一方で、疫学研究に関する倫理指針や健康増進法が施行されたとはいえ、即健診データの活用が容易になるとは思われない。その拒否理由として、条例で認められていない、個人情報保護審査会の検討困難、インフォームドコンセントをとっていない、結果が公表された場合の住民への対応困難、などが現場の声として上げられる。

大島が提起しているように、手続き論に終始することなく、public opinionを重視した、関係市町村や住民への地道な説得が必要かと思われる。

(3) 記録照合における問題点

データ管理のシステムが異なるため、入出力フォーマットの違いや文字化けしていることがある。データ入力の際と変数定義(コード化、個人識別、住所コードなど)に関しては誤解のないように十分な情報交換をする必要がある。

5. 倫理面の配慮

住民の個人情報扱うことになるのでその秘密保持には万全を期するのは当然である。基本健診データについては、データ提供依頼に対して承諾の得られた市町村に限定した解析を行っている。今回紹介した疫学研究の最近のものについては、「疫学研究に関する倫理指針」に従って、「鳥取県における健診データを活用したがん罹患リスクの解明」という研究テーマで、2003年1月30日付けで鳥取大学医学部倫理審査委員会の承認を得ている。また、鳥取県がん登録データに関しても、既に鳥取県個人情報保護審査会において、平成11年10月にがん登録について本人以外から情報を収集することが承認されている。

6. おわりに

今回は、鳥取県における健診データとがん登録データの記録照合による後ろ向きコホート研究への活用例およびその問題点と限界について言及した。

健診データの活用方法には、以上紹介した以外にも沢山あると思われる。

総じて、後ろ向きコホート研究の一環として、制約はあるが健康診断情報を長期継続観察していけば、大規模集団における健診情報項目とがん罹患の関係についての知見・エビデンスは年々蓄積され、より信頼性の高いものとなっていくと考える。

そのためにも、地域がん登録の精度向上をめざした取り組みは益々重要であることは言うまでもない。

文献

1. 陶山昭彦:基本健康診査受診群のがん罹患特性に関する疫学的研究.米子医学雑誌:219-235,1997.
2. 陶山昭彦,岡本幹三,岸本拓治:耐糖能異常者のがん罹患リスクについての検討.厚生省がん助成金「地域がん登録の精度向上

- と活用に関する研究」平成 11 年度報告書：137-139, 1999.
3. 岸本拓治, 尾崎米厚, 岡本幹三: 肺がん検診受診者における喫煙状況別がん罹患に関する研究 . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 12 年度報告書：159-161, 2000.
 4. 岡本幹三, 嘉悦明彦, 尾崎米厚, 岸本拓治: 鳥取県における基本健診受診者のがん罹患に関する研究 血清脂質とがん罹患リスク . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 13 年度報告書：156-159, 2001.
 5. 岸本拓治, 岡本幹三, 尾崎米厚, 嘉悦明彦: 「死の四重奏」と乳がん罹患リスク . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 14 年度報告書：156-159, 2002.
 6. 岡本直幸, 夏井佐代子: 地域がん登録とのレコード・リンケージによる肺がん検診受診者のがん罹患追跡調査、厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 6 年度報告書 .211-216, 1994.
 7. 津熊秀明, 花井彩, 内藤義彦, 飯田稔, 大島明: 検診ファイルと地域がん登録資料との記録照合によるコホート研究 . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 6 年度報告書 217-221, 1994.
 8. 大島明, 内藤義彦, 飯田稔, 津熊秀明, 花井彩, 大塚順子: 検診ファイルと地域がん登録資料との記録照合によるコホート研究 . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 7 年度報告書 . 175-179, 1995.
 9. 内藤義彦, 飯田稔, 津熊秀明, 大島明, 大塚順子: 循環器検診受診者のがん登録との記録照合による追跡研究 . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 8 年度報告書 .168-169, 1996.
 10. 早田みどり, 池田高良: 大腸がんの罹患・死亡に及ぼす検診の影響 検診群と非検診群の比較 . 厚生省がん助成金「地域がん登録の精度向上と活用に関する研究」平成 10 年度報告書 . 177-180, 1998.

Summary

By the record linkage of mass health examination and cancer registry in Tottori prefecture we have carried out the following retrospective cohort studies until now;

1. Epidemiological study of characteristics about cancer incidence of participants in health screening
2. Impaired glucose tolerance and cancer risk
3. Serum lipids and cancer risk
4. Smoking habits and lung cancer risk
5. The deadly quartet and breast cancer risk
6. Hypertension and cancer risk

Above these retrospective cohort studies have shown that impaired glucose tolerance, smoking habits, obesity and having two or more the selected risk factors are significantly associated with incidence of cancer.

On the other hand, the lack of the data regarding eating habits in health examination and the difficulty of obtaining consent of data offer in cities, towns and villages were pointed out as a problem.

In any case, it is important to utilize the record linkage of health examination and cancer registry as a means which reveals evidence of population-based cancer risk in the future.