

B. 研究方法

長崎県がん登録では、県を通じ死亡小票を紙ベースで入手している。入手のタイミングは、死亡から1-2ヶ月遅れとなる。まずそれをがん（脳の良性および性状不詳の腫瘍も含む）の記載のあるものと記載のないものに分け、がんの記載のない死亡小票に関しては、登録データと照合し死亡日を特定している。また、がんの記載のあるものに関しては、原死因か否かを判断し、死因をデータベースに入力している。入力項目は、姓、名、生年月日、死亡年月日、死亡時住所、がん（脳の良性および性状不詳の腫瘍も含む）病名をコード化したもの、死因か否かである。死因か否かに関しては、以下の区分を行っている。

- A：1欄記載で原死因である。
- B：2欄記載だが原死因である。
- C：1欄2欄を問わず疑いとなっている。
- D：2欄記載で原死因でない。
- E：1欄記載だが原死因とならない。

通常、登録室職員がこの判断を行うが、判断に迷う症例は登録室長である医師がこの判断を行っている。年報作成時にはA、Bの悪性腫瘍のみを対象として死亡統計を作成している。

今回、長崎県を通じ2007年と2008年の人口動態テープを入手することができた。

登録室が作成した死亡データと人口動態死亡データの比較検討を行う目的で、両データの照合を行った。人口動態テープには個人を特定する情報である氏名と住所の情報がないために、生年月日と死亡年月日および性別の情報を用いて登録室で作成した死亡データと照合した。

C. 研究結果

表1に示すように、がんあるいは脳の良性または性状不詳腫瘍の記載のある死亡小票の数は登録室では2007年に5193件、2008年には5331件あったのに対し、人口動態テープでは夫々4786件、4909件であった。一致したものは夫々4751件、4886件であった。登録室では県外で死亡した場合でも県内に戸籍があると死亡情報が移送分として含まれてくるため、登録室データのみの442件中65件(14.7%)と445件中56件(12.6%)は死亡時住所が県外分であり、当然、人口動態テープには含まれていなかった。人口動態テープにはあるが、がん登録にはデータの存在しないものが、2007年は35例、2008年は23例認められた。

表1 登録室作成死亡データと人口動態死亡データの比較

	登録室データ	人口動態テープ	同一人	登録室データのみのみ	人口動態テープのみのみ
2007年	5193	4786	4751	442(65)	35
2008年	5331	4909	4886	445(56)	23

()内の数値は死亡時住所が県外のもの

双方にデータのある死亡症例のうち、診断名が一致したものは2007年では4632件、2008年には4747件、一致しなかった症例

は夫々119件(2.5%)、139件(2.8%)であった。

表2 同一人に関する診断名の一致状況

	診断一致	診断不一致
2007年	4632 (97.5%)	119 (2.5%)
2008年	4747 (97.2%)	139 (2.8%)

登録室の死亡データと人口動態テープの死亡データ両方に共通した症例に関して、登録室での死因取り扱い分類を調査した。表3はその結果を示している。2007年、2008年ともに1欄記載で原死因である症例が夫々4537例、4635例と大多数を占めていた。2欄記載で原死因である症例は夫々133例、135例であった。1欄記載であっても疑い例

は死亡統計からは除外している。また、1欄記載であっても、「肝硬変、肝がん」のような記述がされているものについては肝硬変を死因とみなすことでこれまで登録を行ってきた。表3だけを見ると、年間100例前後が長崎県統計では抜けていることになる。

表3 合致例における登録室での死因取り扱い分類

	1欄記載で原死因	2欄記載で原死因	疑診症例	2欄記載で原死因でない	1欄記載で原死因でない
2007年	4537 (95.5%)	136 (2.9%)	15 (0.3%)	58 (1.2%)	5* (0.1%)
2008年	4635 (94.9%)	144 (2.9%)	16 (0.3%)	88 (1.8%)	3* (0.1%)

* 8例とも、1欄アに「肝硬変、肝がん」と書かれており、初筆である肝硬変を死因とみなした。

次に、人口動態テープにはデータがない、つまり、死因として取り上げられていない症例で、登録室でのみ確認された症例について、登録室での死因取り扱い分類を調査した。表4に示すように、死亡時住所が県外の者が2007年14.7%、2008年12.6%含まれていた。また、2欄記載で原死因ではない症例が2007年は279例(63.1%)、2008年は266例(59.8%)と大半を占めるが、原死因とみなしたものが2007年94例

(21.2%)、2008年115例(25.8%)認められた。表3のデータと併せると、1欄記載で原死因である症例が2007年は4546例、2008年は4656例、2欄記載で原死因である症例は夫々221例、238例となり、合計すると2007年には4767例、2008年には4894例が原死因として計上されたことになる。国の死亡統計との差は2007年が19例、2008年は15例であった。

表4 人口動態テープになく、登録室でのみ確認された症例の死因取り扱い分類

	1欄記載で 原死因	2欄記載で 原死因	疑診症例	2欄記載で 原死因でない	1欄記載で 原死因でない	県外死亡
2007年	9(2)	85(17)	3	279(33)	1(1)	65
	2.0%	19.2%	0.7%	63.1%	0.2%	14.7%
2008年	21(5)	94(13)	8(2)	266(38)	0	56
	4.7%	21.1%	1.8%	59.8%	0.0%	12.6%

()内の数値は性状不詳の腫瘍を示す

最後に、登録室データに無く、人口動態テープでのみ確認された症例について分析を行った。原票が確認できず、恐らく入手されていないと思われる症例が2007年に29例、2008年に14例認められた。原票が見つかり、脳の良性腫瘍であったものが1例

見つかったほかは、脳以外の良性腫瘍或いは人口動態テープのコードミスと思われた。表6に原死因コードと原票記載死因とが矛盾する症例一覧を示した。

表5 登録室データになく、人口動態テープのみで確認された症例の死因取り扱い分類

	原票あり、 良性腫瘍	原票あり、原死因コードと 原票記載死因とが矛盾	原票あり、登録室 の見落とし	原票なし
2007年	2 (5.7%)	3 (8.6%)	1 (2.9%)	29 (82.9%)
2008年	4 (17.4%)	5 (21.7%)	0	14 (60.9%)

表6 人口動態テープの原死因コードと原票記載死因とが矛盾する症例一覧

ICDコード	疾患名
C64	腎臓症
C169	(工欄)十二指腸断端縫合不全
C349	てんかん重積発作
C169	転落(外傷性肝破裂)
C169	食道損傷
C349	大動脈瘤破裂
C220	多発性肝症
D376	胆道ホリブ

(倫理面への配慮)

本研究は、登録作業の一環として行ったものであり、通常の疫学研究とは異なる。

D. 考察

わが国における地域がん登録の精度は非常に低く、欧米諸国との間に大きな隔りがある。がん登録に関する法的整備が充分でないことが、がん登録の精度向上を困難にしていることは否めない。一方、死亡統計は法律の裏づけがあり、きわめて正確であると考えられている。今回、初めて入手できた人口動態テープと登録室で作成した死亡データとの比較検討を行い、若干の相違を認めた。わずか2年間の症例検討であるが、国が作成した死亡統計によると、長崎県では2年間で9695例のがん或いは良性または性状不詳の腫瘍死が認められていた。一方、長崎県がん登録が作成した死亡統計では、9661例のがん或いは良性または性状不詳の腫瘍死が認められていた。国の正式な統計の方が2年間で34例多かったことになる。今回は詳細な診断の食い違い等についての検討は行っていない。

原死因の特定に若干の違いがあるのは想定範囲内であるが、今回の検討で、死亡小票の入手に漏れがあることが判明した。また、また、登録室では明らかに死因としてカウントしている症例が、国の統計から除外されていることより、人口動態テープの死因統計にも漏れがある可能性が示唆された。死亡統計を得るのみでなく、罹患統計にも利用されることを考慮すると、今後、診断不一致例に関する検討も含め、内容を詳細に検討する必要があると考えられた。また、入手できていない死亡小票の入手を計画している。

E. 結論

死亡データは、前述したように、死亡統計に用いるのは勿論、生存確認或いは罹患データの補完にも利用される。もれなく情報収集する仕組み造りが肝心と考えられた。

F. 健康危険情報 特になし

G. 研究発表

1. 論文発表

- 1) Kokichi Arisawa, Midori Soda, Masahito Ono, Hirokazu Uemura, Mineyoshi Hiyoshi, Akihiko Suyama. Trends of incidence rate of adult T-cell leukemia/lymphoma in an HTLV-1 endemic area in Japan. *Int. J. Cancer*. 125, 737-738, 2009
- 2) Daisuke Haruta, Kiyotaka Matsuo, Shinichiro Ichimaru, Midori Soda, Ayumi Hida, Nobuko Sera, Misa Imaizumi, Eiji Nakashima, Shinji Seto, Masazumi Akahoshi. Men with Brugada-like electrocardiogram have higher risk of prostate cancer. *Circ J*. 73, 63-68, 2009
- 3) Wan-Ling Hsu, Midori Soda, Nobuo Nishi, Dale Preston, Sachiyo Funamoto, Masao Tomonaga, Masako Iwanaga, Akihiko Suyama, Fumiyoshi Kasagi. Leukemia, Lymphoma, and Multiple Myeloma Incidence in the LSS Cohort - 1950-2001. *Radiation Health Risk Sciences*. Tokyo, Japan: Springer; 2009. p. 69-73
- 4) Suyama A, Izumi S, Koyama K, Sakata R, Nishi N, Soda M, Grant E, Shimizu Y, Furukawa K, Cullings HM, Kasagi F, Kodama K. The offspring of atomic bomb survivors:

Cancer and non-cancer mortality and cancer incidence. *Radiation Health Risk Sciences*. Tokyo, Japan: Springer;. 57-62, 2009

5) S Miura, M Nakashima, H Kondo, M Ito, S Meirmanov, T Hayashi, M Soda, I Sekine. Significance of oncogene amplifications in breast cancer in Atomic Bomb survivors: associatin with radiation exposure and histological grade. *Radiation Health Risk Sciences*. Tokyo, Japan: Springer;. 285-294, 2009

6) David B. Richardson, Hiromi Sugiyama, Steve Wing, Ritsu Sakata, Eric Grant, Yukiko Shimizu, Nobuo Nishi, Susan Geyer, Midori Soda, Akihiko Suyama, Fumiyoshi Kasagi, Kazunori Kodama. Positive Associations Between Ionizing Radiation and Lymphoma Mortality Among Men. *Am J Epidemiol*. 169, 969-976, 2009

7) David Richardson,a,1 Hiromi Sugiyama,b Nobuo Nishi,b Ritsu Sakata,b Yukiko Shimizu,b Eric J. Grant,b Midori Soda,d Wan-Ling Hsu,c Akihiko Suyama,d Kazunori Kodamae and Fumiyoshi Kasagib. Ionizing Radiation and Leukemia Mortality among Japanese Atomic Bomb Survivors, 1950–2000 *RADIATION RESEARCH* 172, 368-382, 2009

8) Yoshito Koga, Masako Iwanaga, Midori Soda, Naoko Inokuchi, Daisuke Sasaki, Hiroo Hasegawa, Katsunori Yanagihara, Kazunari Yamaguchi, Shimeru Kamihira, Yasuaki Yamada. Trend in HTLV-1 Prevalence and Incidence of Adult T-Cell Leukemia /

Lymphoma in Nagasaki, Japan. *J Med Virol*. 82, 668-674, 2010

2. 学会発表

1) Trend of Adult T-cell Leukemia / Lymphoma Incidence in an HTLV-1 endemic areas in Nagasaki Japan using the Joinpoint Analysis. Midori Soda, Masako Iwanaga, Yasushi Miyazaki, Yasuaki Yamada, Kokichi Arisawa, Ichiro Senkine, Akihiko Suyama. 31th Annual Meeting of the International Association of Cancer Registries, 3-5 June 2009, New Orleans, USA

H. 知的所有権の取得状況

1. 特許取得 なし
2. 実用新案特許 なし
3. その他 なし

がん罹患の動向分析

分担研究者 加茂 憲一 札幌医科大学 医療人育成センター 数学・情報科学講座
分担研究者 雑賀 公美子 国立がんセンター がん対策情報センター

研究要旨

日本において報告される全国がん罹患数は、地域がん登録データを用いた推計値である。推計に用いるデータは都道府県レベルで10～15地域であり、登録の完全性に関する基準を充足する地域に限られる。しかし、この基準をみたす地域においても完全性が十分に高いわけではない。また、推計に用いられる手法においても登録の完全性に関する補正は行われないので、全国推計値は過小評価されている可能性が極めて高い。登録の完全性について補正する方法として、回帰モデルを用いる手法が知られている。今回、ロジスティック回帰モデルを用いた推定法に基づく罹患数とその信頼区間の構成について報告する。実際に、2004年罹患数をロジスティック回帰モデルにより推計した。結果については報告値との比較を行い、過小評価の定量的な評価を行った。

A. 研究目的

現在、国立がんセンターより毎年公表されている全国がん罹患数は、都道府県規模で10～15地域の地域がん登録に基づく罹患データおよび人口動態統計に基づく死亡データによる推計値である。推計には日本固有の方法が採られており、データが内包する様々な問題点を可能な限りカバーするように構成されている。具体的には、以下の手順により推計が行われている：

- (1) 男女計全がんに対して、IM比1.5以上かつ、DCN割合30%未満またはDCO割合25%未満を充足する地域を選定する。
- (2) これら地域における、性・年齢階級別の罹患率および死亡率を算出する。
- (3) (2)で算出した罹患率および死亡率に全国人口を乗じ、全国罹患数および死亡数の推計値を得る。
- (4) (3)で得られた全国死亡数推計値と、人口動態統計における全国死亡数の比を、罹患推計における補正係数と定義する。
- (5) (3)で得られた暫定的な全国罹患数推計値に、(4)で得られた補正係数を乗じ、最終的な全国がん罹患数推計値を得る。

この手法は、現在において日本の特性を最大限に活用した推計方法であり、様々な利点を有する。まず、日本の地域がん登録における登録の完全性の問題については、手順(1)において完全性の低い地域を除外することにより全国推計値に完全性の変動が与える影響を排除している。また、日本においては人口動態統計により完全な死亡データが得られることから、手順(4)による補正係数が意味を持つ。この係数を用いることにより、推計に用いる地域の全国に対する偏りが補正される。以上のような利点が存在するが、この手法で完全というわけではなく、更なる手法の改善が望まれている。

その中でも特に、本研究では登録の完全性に着目する。現行の罹患推計においては、手順(1)により完全性の高い地域が選られている。しかし、この基準は国際的な視点では決して厳しいものではない。登録の完全性に関する更なる補正を行うには、基準を厳しくするか、推定の過程における何らかの手法的な改良が必要である。Kamo et al, (2007), Jpn. J. Clin. Oncol.において、DCN割合を説明変数、IM比を被説明変数とする非線形回帰モデルを

用いて、DCN割合が0%におけるIM比を推定することにより、登録の完全な状態での全国がん罹患数を予測する手法が紹介されている。その後、同様の考え方に基づくロジスティック回帰モデルを用いる手法が提案された。今回の報告書では、ロジスティック回帰モデルにおける最尤推定量の漸近正規性に着目し、登録の完全性に関する補正を行った全国罹患数および、その信頼区間を構成する。更に、最新の2004年データを用いた解析結果を報告し、報告値との比較検討を行う。

B. 研究方法

登録の完全性に関する補正を施した全国がん罹患数を推定するために、DCN割合を説明変数とするロジスティック回帰モデルを用いる。地域番号 i ($i = 1, 2, \dots, N$) におけるがん罹患数を n_i 、がん死亡数を m_i とし、罹患から死亡が発生する確率を p_i とする。つまり、 p_i はMI比 (IM比の逆数) を表す。この p_i はDCN割合 x_i の影響を受け変化することに注意を要する。死亡数 m_i は、 n_i と p_i をパラメータとする二項分布に従うとする。つまり

$$m_i \sim \text{Bin}(n_i, p_i)$$

である。MI比調整パラメータを α 、DCN割合が0%におけるIM比 (今後「真のIM比」と呼ぶ) を α_0 とする。ここで $\alpha = (\text{登録群におけるMI比}) \div (\text{未登録群におけるMI比})$ で定義する。具体的には、登録群におけるがん死亡率が、未登録群より高ければ、 α は1より大きくなる。またDCN割合が0%とは、登録が完全な状態と考えられるので、 α は完全登録の下でのIM比と考えられる。これらのパラメータを用いると p_i は次の2次方程式の解のうち大きい方となる：

$$\alpha p_i^2 + (\alpha x_i - \alpha_0 x_i - 1)p_i + (1 - \alpha_0)x_i = 0.$$

この方程式の導出は、罹患を「既登録か、未登録か」と「がん死亡か、それ以外か」の観点での4群分割を用いて、DCN割合・MI比・未

知パラメータを関連付けることにより得られる。一方、MI比調整パラメータが1ならば、 p_i は x_i に関する一次式として、次で簡略に表される：

$$p_i = \alpha^{-1} + \alpha^{-1}(\alpha - 1)x_i.$$

未知パラメータについては $\alpha > 0$ かつ $\alpha > 1$ であることに注目し、 $\alpha_1 = \log \alpha$ 、 $\alpha_2 = \log(\alpha - 1)$ と変数変換を行い、 $\alpha = (\alpha_1, \alpha_2)'$ と設定しなおす。そして

$$m_i \sim \text{Bin}(n_i, p_i(\alpha))$$

とするロジスティック回帰モデルを構成する。

最尤推定量には漸近正規性があることが知られている。具体的には、未知パラメータベクトル α の推定量 $\hat{\alpha}$ が漸近的に

$$N^{1/2} (\hat{\alpha} - \alpha_0) \rightarrow N(0, I(\alpha_0)^{-1})$$

であることを用いて、パラメータの信頼区間を構成する。ここで α_0 は α の真値であり、 I はフィッシャー情報行列である。

次に、MI比調整パラメータ=1の場合と、 $\neq 1$ の場合、どちらのモデルを用いるのが適切かを決定する必要がある。その際には交差検証法 (Cross-validation) を用いる。すなわち

$$CV = N^{-1} \sum (m_i - n_i p_{[-i]})^2$$

を算出し、この値が小さいモデルを適切と判断する。ここで $p_{[-i]}$ は i 番目のジャックナイフ推定量である。CV値は、あるサンプルと、そのサンプル以外のデータを用いた予測値の乖離を表す情報量規準であり、この値が小さいモデルは、予測の観点から優れたモデルであるといえる。

最後に、推定された真のIM比に全国死亡数を乗じることにより、全国罹患数の推定値を得る。ここで、日本における死亡データは、人口動態統計により完全なものが得られていることが前提となっている。

(倫理面への配慮)

今回の解析には、倫理面に配慮を要するデータは用いていない。

C. 研究結果

2004年全国がん罹患数推定に用いられたのは、14府県（宮城県、山形県、千葉県、神奈川県、新潟県、福井県、滋賀県、大阪府、鳥取県、岡山県、広島県、佐賀県、長崎県、熊本県）のデータである。今回、同じ地域のデータを用いて、ロジスティック回帰モデルに基づく罹患数全国推定を行った。採択されたモデルおよび未知パラメータに関する結果を表1に示す。

表1
選択されたモデルと未知パラメータの推定値

性別	部位	モデル	真のIM比	真のMI比の95%信頼区間
男性	全がん	1	2.667	(2.639, 2.695)
	胃	1	3.018	(2.942, 3.098)
	結腸	1	4.061	(3.877, 4.257)
	直腸	2	4.254	(3.086, 6.078)
	肝臓	1	1.378	(1.354, 1.403)
	肺	1	1.660	(1.634, 1.688)
	前立腺	1	8.698	(8.064, 9.389)
女性	全がん	1	3.121	(3.077, 3.167)
	胃	1	2.747	(2.652, 2.848)
	結腸	1	3.192	(3.057, 3.336)
	直腸	2	5.106	(3.145, 8.867)
	肝臓	1	1.480	(1.435, 1.529)
	肺	1	1.849	(1.794, 1.907)
	乳房	1	6.200	(5.916, 6.501)

モデルは、「1」が $\square=1$ のモデル、「2」が $\square \neq 1$ のモデルを表す。真のIM比は、選択されたモデルのもとでの \square の推定値である。

次に、本モデルにより推定した全国がん罹患数を表2に示す。表1・2共に部位に関しては全がんに加えて代表的な5部位（胃、結腸、直腸、肝臓、肺）と、男性の前立腺、女性の

乳房に関する結果を示す。

表2 全国罹患数の推定結果
男性

部位	推定値 (95%信頼区間)	報告値	差（推定値－報告値）
全がん	514987 (509580, 520394)	372913	142074
胃	99144 (96648, 101772)	73950	25194
肺	72909 (71943, 74139)	55984	16925
結腸	54032 (51583, 56639)	35657	18375
直腸	36844 (26728, 52642)	21421	15423
肝臓	32274 (31501, 32860)	28172	4102
前立腺	76890 (71286, 82999)	39321	37569

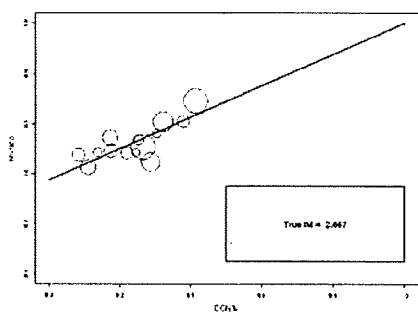
女性

部位	推定値 (95%信頼区間)	報告値	差（推定値－報告値）
全がん	397185 (391585, 403039)	275578	121607
胃	48652 (46970, 50441)	35822	12830
肺	29586 (28706, 30514)	24122	5464
結腸	42029 (40252, 43925)	29070	12959
直腸	26582 (16373, 46162)	11907	14675
肝臓	16412 (15913, 16955)	13343	3069
乳房	65249 (62260, 68417)	50549	14700

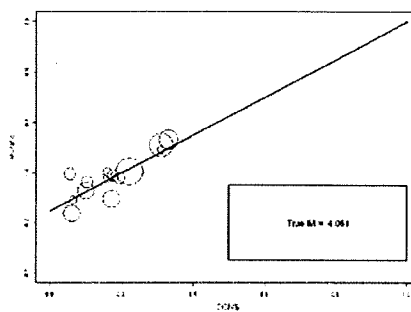
「推定値」はロジスティック回帰モデルに基づく手法による推定値、「報告値」は国立がんセンターより報告されている数値である。これらの差が登録漏れ数と考えられる。

次に、推定に用いた地域のDCN割合－MI比プロットと、推定された回帰曲線を図1に示す。横軸はDCN割合、縦軸はMI比を表し、各プロットの面積は、罹患数の報告値に比例する。

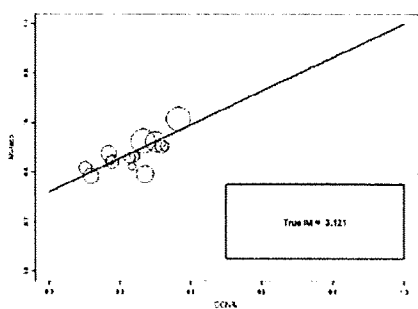
図1 DCN割合—MI比プロットと回帰曲線
全がん（男性）



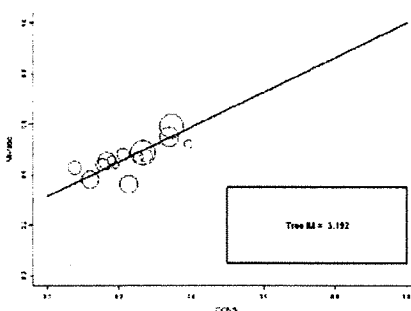
結腸（男性）



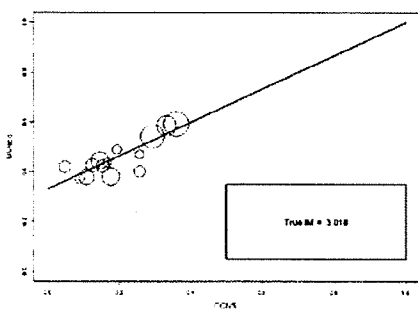
全がん（女性）



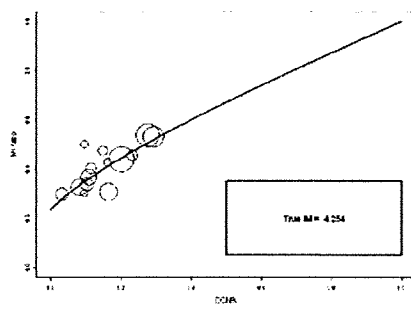
結腸（女性）



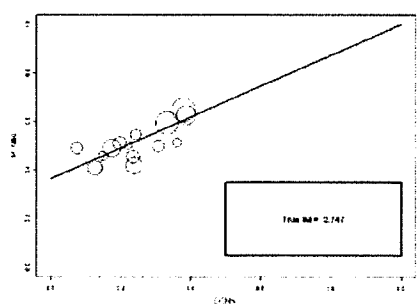
胃（男性）



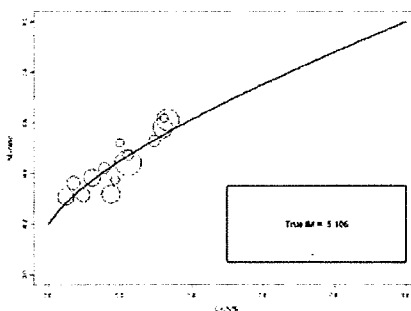
直腸（男性）



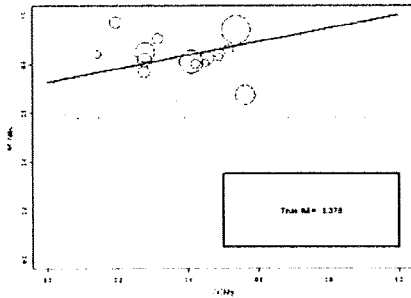
胃（女性）



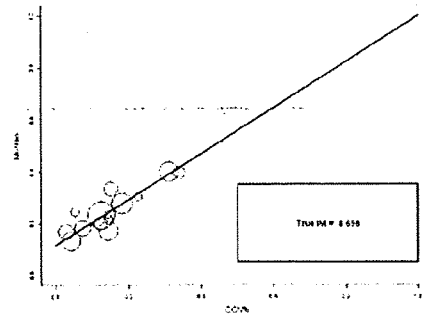
直腸（女性）



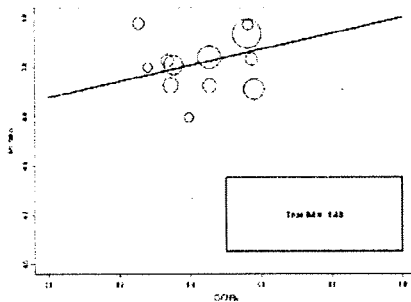
肝臓 (男性)



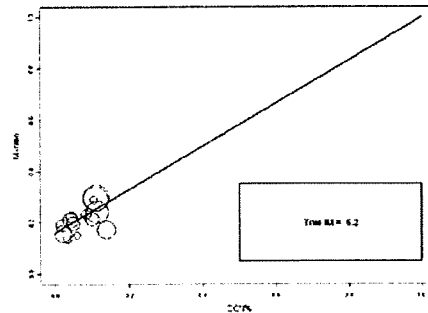
前立腺



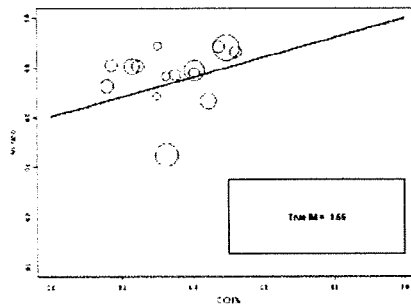
肝臓 (女性)



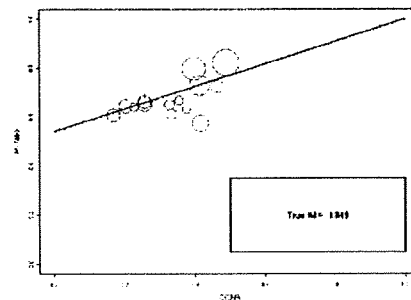
乳房



肺 (男性)



肺 (女性)



横軸はDCN割合、縦軸はMI比を表す。また、各プロットの面積は、人口に比例する。

D. 考察

考察対象の部位の中では、直腸がんについては非線形の複雑なモデルが選ばれたが、その他については単純な線形モデルが選ばれた。全国罹患数を推定する手法に関する先行研究の多くで $\alpha=1$ という仮定がおかれている。今回の結果より、多くの部位においてこの仮定は妥当であるといえるが、結腸がんにおいては複雑なモデルが選ばれたことから、MI比調整パラメータの必要性についての議論が不必要とは限らない。

今回の解析においては、現行の報告値との比較を正確に行うため、2004年罹患推定に用いられた13地域のみを解析に用いた。しかし、この手法は回帰モデルに基づいており、様々なDCN割合に対する観測値がある方が望ましいと考えられる。つまり登録の完全性に関する

基準をクリアしていない地域データを用いても理論上問題はない。登録の完全性が低い地域においても、その地域は低いなりの情報を持っていることが予想される。統計解析においては、サンプル数は多い方が望ましいことから、収集した全ての地域のデータを用いられる点も本モデルの利点である。

回帰曲線のあてはまりを見ると、良い部位と悪い部位が混在する。特に肝臓・肺においては直線モデルが選ばれているが、見た目のあてはまりは決して良くない。これは、同じDCN割合に対するMI比のパラツキが大きいことに起因していると考えられる。想定した「MI比はDCN割合で説明できる」というモデルでは説明できない要因（例えば地域効果や、罹患・死亡の発生年のずれ）が存在することを予想させる結果である。しかし全体的には、MI比とDCN割合は関連性が高く、回帰曲線のあてはまりも良いといえよう。

E. 結論

今回、登録の完全性を補正した全国がん罹患数の推計方法を提案した。この方法は、DCN割合が0%におけるIM比を、DCN割合を説明変数、MI比を被説明変数とする回帰モデルを用いて推定するというアイデアに基づいている。

今回、最新の罹患報告である2004年について解析を行った。同年の罹患数報告値は65万人であるのに対し、本報告書で提案する方法では91万人という推定結果が得られた。本結果がある程度正しいと仮定すると、26万人過小評価されている可能性があると考えられる。言い換えると、30%弱の過小評価の可能性はある。このように、本報告書で提案する手法は全国推計値の予測を与えるだけでなく、登録漏れを定量的に計る指標の1つにもなる。

G. 研究発表

1. 論文発表

- 1) T. Matsuda, T. Marugame, K. Kamo,

K. Katanoda, W. Ajiki, T. Sobue, Cancer incidence and incidence rates in Japan in 2003: based on data from 13 Population-based cancer registries in the monitoring of cancer incidence in Japan (MCIJ) project, Jpn J Clin Oncol, 39(12): 850-858, 2009.

2. 学会発表

- 1) K. Saika, K. Kamo, W. Ajiki, T. Sobue, National cancer incidence estimation by Poisson regression in Japan, 31th Annual Scientific Congress and Meeting of the International Association of Cancer Registries, アメリカ, 2009.
- 2) 雑賀公美子, 加茂憲一, 味木和喜子, 祖父江友孝, 自己回帰モデルを用いたがん死亡率短期予測の可能性, 日本がん予防大会2009愛知, 愛知, 2009.
- 3) 加茂憲一, 片野田耕太, 松田智大, 丸亀知美, 味木和喜子, 祖父江友孝, Lifetime probability of developing or dying of cancer in 2003, 日本癌学会, 横浜, 2009.

F. 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 |

厚生労働科学研究補助金（第3次対がん総合戦略研究事業）
分担研究報告書

国際がん研究機関（IARC）と日本の全国罹患数推計方法の比較

研究分担者 雑賀公美子¹⁾、加茂憲一²⁾

研究協力者 Hai-Rim Shin³⁾、Jacques Ferlay³⁾、Graham Byrnes³⁾、中村 隆⁴⁾

1) 国立がんセンター がん対策情報センター がん情報・統計部

2) 札幌医科大学 医療人育成センター 数学・情報科学講座

3) International Agency for Research on Cancer

4) 情報・システム研究機構 統計数理研究所

研究要旨

日本で公表されているがんの部位、年齢階級別罹患数および罹患率は一部の地域がん登録の情報を用いた推計値であり、推計方法は日本独自の方法で行われている。また、国際がん研究機関（IARC）では、全世界のがんの罹患、死亡、有病に関する情報を GLOBOCAN として公表しており、ここにも日本のがん罹患数の推計値が報告されている。本研究では、まず分析対象データを同一にして日本および IARC で用いられている方法による全国罹患推計を比較しその違いを明らかにした。さらに、推計に用いる地域がん登録の実測罹患率と推計罹患率との差が推計に用いる地域によってどの程度異なるのかを明らかにした。

がん罹患数の全国推計値については、同一データに基づけば日本法によるものも IARC 法によるものも、罹患率の高い部位では大きな差はないことが明らかになった。しかし、甲状腺がんなどの罹患数が少ない部位では、推計罹患数、年齢調整罹患率に違いがみられた。より精度の高い全国罹患推計を行うためのモデルを構築することが大切であり、現在の日本法および IARC 法を元に、年齢変数の扱い方や、モデル選択の方法、人口動態統計により完全に把握されている死亡情報の扱いなどを考慮した、さらなるモデルの構築が必要である。

A. 研究目的

日本で公表されているがんの部位別年齢階級別罹患数および罹患率は一部の地域がん登録の情報を用いた推計値である。この推計は日本独自の方法で行われており、その方法および推計値はがん対策情報センター Web ページにおいて毎年公表されている。一方、国際がん研究機関（International

Agency for Research on Cancer, IARC）では、全世界のがんの罹患、死亡、有病に関する情報を GLOBOCAN として公表しており（最新は GLOBOCAN2002）、ここにも日本のがん罹患数の推計値が報告されている。日本と IARC の罹患の公表値の違いは、推計方法の違いによるだけでなく、推計に用いられた地域がん登録の罹患データの違

いによるものもある。本研究では、同一の地域がん登録の罹患データ（2002-2004年の3年分）を用いることにより、日本独自の方法とIARCで用いられている方法による2004年の全国罹患推計値を比較し、その違いを明らかにする。さらに、推計に用いる地域がん登録の実測罹患率と推計罹患率との差が推計に用いる地域によってどの程度異なるのかを明らかにした。

B. 研究方法

1. 資料・対象

地域がん登録で集計される地域別の罹患数は、登録の精度が安定している地域のもののみを用いることとした。精度の指標は、罹患-死亡比 (incidence-mortality ratio, IM比)、死亡票によりはじめて登録されたものの割合 (death certificate notification, DCN 割合)、死亡票のみによる登録の割合 (death certificate only, DCO 割合) とし、日本が罹患数の全国推計を行う際に用いている基準に従った。具体的には男女計全がん(C00-C96、国際疾病分類第10版による)に対してIM比1.5以上、かつDCN割合30%未満またはDCO割合25%未満を満たすことである。

本研究で用いた地域のデータは、結果的に、宮城、山形、神奈川、新潟、福井、滋賀、大阪、鳥取、岡山、佐賀および長崎(以上11地域の2002-2004年の3年分)、千葉および広島(以上2地域の2003-2004年の2年分)、熊本(2004年の1年分)である。

全国および都道府県の死亡数は、人口動態統計¹⁾による2002-2004年の年別、性・年齢5歳階級別死亡数を用いた。全国および都道府県人口は、5年ごとに実施されて

いる国勢調査人口および国勢調査時の人口を基準に、推計した人口²⁾を用いた。この人口は、地域がん登録集計用として国立がんセンターで推計しているものである。

2. 日本で用いられている全国罹患数推計方法(日本法)³⁾

日本法は以下の手順で行われている。

- (1) 各地域の性別年齢階級別の平均罹患率および死亡率を算出する。年齢階級区分は、0-4、5-9、...、85歳以上の18階級である。
- (2) (1)で算出した平均罹患率および死亡率に全国人口を乗じ、全国の推計罹患数および推計死亡数とする。
- (3) (2)で推計した全国推計死亡数と人口動態統計で報告されている全国死亡数の比を、補正係数とする。
- (4) (2)で推計した全国推計罹患数に(3)の補正係数を乗じ、全国がん罹患数推計値とする。

3. IARCで用いられている全国罹患数推計方法(IARC法)⁴⁾

IARC法は以下の手順で行われている。

- (1) 登録地域別の罹患数(C_i : 地域*i*の罹患数)を用い、観測された全国の性別年齢階級別罹患数(C)を算出する。

$$C = \sum_i C_i$$

- (2) (1)で算出した性別年齢階級別罹患数(C)、登録地域別の罹患数(C_i)および登録地域別の死亡数(D_i : 地域*i*の死亡数)を用い、ウェイトを $1/\sqrt{P_i}$ (P_i は地域*i*の人口)とした重み付けした全

国の性別年齢階級別死亡数(D)を算出する。

$$D \approx C \frac{\sum_i (D_i / \sqrt{P_i})}{\sum_i (C_i / \sqrt{P_i})}$$

- (3) 全国の性別年齢階級別罹患数 (C) がポアソン分布に従うと仮定。オフセットを(2)で算出した重み付き全国性別年齢階級別死亡数(D)とし、性別と年齢を説明変数としたモデルを設定する。年齢階級は、0-44、45-49、50-54、55-59、60-64、65-69、70-74、75-79、80-84、85歳以上の10階級である。若年の45歳未満が1階級として扱われていることに注意。
- (4) (3)のポアソンモデルのモデル選択では、罹患-死亡比 (IM 比) に対してもっともあてはまりのよいモデルをデビアンスの改善度を基準に決定する。より具体的には、まず性別変数の投入の有無を決定し、その後年齢変数の次数 (1次から5次) を決定する。
- (5) (4)で選択されたモデルのパラメータ推定値を用い、性別年齢階級別の IM 比を算出する。
- (6) (5)で算出した性別年齢階級別の IM 比に性別年齢階級別の全国死亡数を乗じ、全国罹患数推計値とする。

4. 推計に用いるデータによる推計誤差

日本法においては、全国推計値に一定の精度基準を満たした地域がん登録の罹患情報を用いるため、年によって用いられる地域が異なる (2002年11地域、2003年13地域、2004年14地域)。日本法とIARC法において、推計に用いる地域がん登録の実

測罹患率と推計罹患率との差が推計に用いる地域によってどの程度異なるのかを明らかにするため、交差検証法 (クロスバリデーション) を用いた。

2004年の全国推計値を算出するために用いた2002-2004の3年の地域がん登録のデータ38セットを用い、3または4データセットから得られる観測罹患率を、残りの35または34データセットから日本法およびIARC法で推計した推計罹患率の比の算出を11回繰り返した。この11個の比の分布を誤差分布として両方法を部位別に比較した。

C. 研究結果

1. 推計罹患数、年齢調整罹患率

表1にIARC法と日本法のそれぞれで推計した全国罹患数と年齢調整罹患率を示した。年齢調整罹患率の高い群 (グループ1)、中程度の群 (グループ2)、低い群 (グループ3) の3群に分けて検討した。

男女とも、日本法による全国推計値の方が高い傾向があった。日本法による推計値が低かった部位は、男性の肝臓、口腔・咽頭および甲状腺のみである。

日本法とIARC法による推計値の違いをみるために、地域がん登録の精度指標の1つであり、IARC法ではモデル化の対象となっているIM比を指標として、両方法によるIM比の比を算出した。IM比については、罹患数/死亡数の比 (以降「数のIM比の比」とする) と、年齢調整罹患率/死亡率の比 (以降「率のIM比の比」とする) の2種類のIM比について検討した。

数および率のIM比の比をみると、男女ともグループ1およびグループ2ではほと

んど差が見られなかった。グループ 3 では男性の甲状腺がんにおいて、数の IM 比の比が 0.72、率の IM 比の比が 0.49 と大きく異なっていた。男性の精巣がんにおいては数の IM 比の比には大きな差はみられないものの、率の IM 比の比が 0.77 であった。グループ 3 の女性では、率の IM 比の比では差があまりみられないものの、喉頭がん、甲状腺がんおよびホジキン病で数の IM 比の比で若干の違いが観察された。

2. 年齢階級別推計罹患率

図 1-1～図 1-3 にグループ別の年齢階級別推計罹患率を示した。グループ 1 およびグループ 2 では男性の直腸、肺および食道において高齢部分で方法による推計値の違いがみられるものの、傾向は類似している。グループ 3 は罹患率が低いため比較は難しいが、表 1 に示した、率の IM 比の比が大きかった男性の精巣と甲状腺においては若年部分においても方法による違いがみられた。

3. 推計に用いるデータによる推計誤差

図 2-1～図 2-3 にグループ別のクロスバリデーション法による観測罹患率と推計罹患率の誤差分布を示した。グループ 1 では食道および肝臓において、男女とも IARC 法より日本法での誤差が大きい。グループ 2 では子宮（体部）で IARC 法の誤差が大きかった。グループ 3 では、女性の喉頭や甲状腺およびホジキン病などは、どちらの方法においても誤差が大きい。方法による違いについては、精巣や甲状腺は IARC 法の誤差が大きく、男性のホジキン病は日本法の誤差が大きい。

D. 考察

日本法と IARC 法の主な違いは、年齢階級の区分とモデルの中での年齢変数の扱い方、モデル選択の有無である。

年齢階級の区分については、日本法では、0歳から85歳未満まですべて5歳階級で扱われており、85歳以上がまるめられている。これに対して、IARC 法では若年の年齢 0歳以上45歳未満が1つの年齢階級とされている。年齢変数の扱いについては、日本法では階級別に推計が行われている。これに対し、IARC 法では年齢は階級同士が関連したものとして扱われている。

モデル選択については、日本法では、性と年齢が常にそれぞれ別々に扱われている。これに対し、IARC 法では性と年齢変数を対象にデビアンズに基づき、当てはまりのよいモデルを決定するためのモデル選択を実施している。

推定された年齢階級別の罹患死亡比（IM 比）をみると、日本法では年齢階級別に扱うため、特に罹患率の低い部位においては推計された IM 比にばらつきが見られる。これに対し、IARC 法では、性・年齢を説明変数として IM 比をモデル化しているため年齢に対して一定の傾向の IM 比が推定される。

本研究により、日本法と IARC 法の全国罹患推計値の比較および用いるデータによる推計の誤差について検討を行った結果、罹患率が高い部位については、年齢階級別では異なる部分もあるが、全体の推計値に大きな違いはないことが明らかとなった。用いる地域の罹患データの違いによる推計誤差は食道、肝臓の日本法で大きく、子宮

(体部)ではIARC法で大きいことが明らかになった。一方、罹患率の低い部位では、精巣や甲状腺で年齢階級別の推計値が推計方法によって異なっていた。用いる地域がん登録データによる推計誤差も罹患率の多い部位に比べて、方法を問わず大きかった。

これらのことから、日本法とIARC法の改善点を考えると、日本法においては、年齢階級別で罹患推計を行っていることで、特に罹患数の少ない部位、少ない年齢階級において推計が不安定であるため、年齢階級同士が関連したものとして扱うモデルを考える必要がある。IARC法においては45歳未満の年齢階級の区分を細分化することで、若年者に罹患の多い子宮や乳房の罹患推計の精度を上げることが期待できる。一方、死亡数をオフセットに用いたポアソンモデルを用いているため、死亡数がゼロの情報には考慮されない。よって、罹患率の低い部位や、生存率が高い部位で罹患数が過小推計されることとなる。

E. 結論

地域がん登録データを用いた、がん罹患数および罹患率の全国推計値については、IARC法も日本法も罹患率の高い部位では大きな差はないことが明らかになった。実際に報告されている国立がんセンターの推計値とIARCの推計値の相違は、使っているがん登録地域や時期によるところが大きい。

しかし、甲状腺がんなど罹患数が少ない部位では、推計罹患数、年齢調整罹患率に違いがみられる。

より現実に則した全国罹患推計を行うためのモデルを構築することが大切であり、

現在のIARC法、日本法を元に、モデルの中でも年齢変数の扱い方や、モデル選択の方法、人口動態統計により完全に把握されている死亡情報の扱いなどを考慮した、さらなるモデルの構築が必要である。

F. 研究発表

1. 論文発表

なし

2. 学会発表

- 1) Saika K, Kamo K, Ajiki W, Sobue T. National cancer incidence estimation by Poisson regression. 31th Annual Scientific Congress and Meeting of the International Association of Cancer Registries, 2009年6月.
- 2) 雑賀公美子, 加茂憲一, 味木和喜子, 祖父江友孝. 自己回帰モデルを用いたがん死亡率短期予測の可能性. がん予防大会 2009, 愛知, 2009年6月.

G. 知的財産権の出願・登録状況

特になし

参考文献

- 1) 厚生労働省大臣官房統計情報部. 人口動態統計, 平成14年~平成16年上巻, 2004~2006.
- 2) 国立がんセンター がん対策情報センター がん情報・統計部.
<http://ncrp.ncc.go.jp/pop.html>, 地域がん登録の技術支援のページ, 2010.
- 3) Matsuda T, Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T. Cancer incidence and incidence rates in Japan in 2003: based on data from 13

population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. *Japanese Journal of Clinical Oncology*, 39(12): 850-8, 2009.

- 4) F. Bray, R. Sankila, J. Ferlay, D.M. Parkin. Estimates of cancer incidence and mortality in Europe in 1995. *European Journal of Cancer*, 38: 99-166, 2002.

表 1. 推計全国罹患数および年齢調整罹患率

	Male						Female					
	Japan method		IARC method		ratio of I/M ratio		Japan method		IARC method		ratio of I/M ratio	
	Number of cases	ASR	Number of cases	ASR	Number of cases	ASR	Number of cases	ASR	Number of cases	ASR	Number of cases	ASR
Group1: Sites with Age-standardized incidence rate large												
Oesophagus	14662	11.37	14503	11.30	1.01	1.01	2728	1.61	2503	0.95	2503	0.95
Stomach	74651	56.81	71086	53.90	1.05	1.05	36475	21.93	35034	9.19	35034	9.19
Colon	36168	27.19	35190	26.88	1.03	1.01	29706	17.28	29017	6.51	29017	6.51
Rectum	23004	18.45	22013	18.03	1.05	1.02	12781	8.36	12526	2.92	12526	2.92
Liver and intragepatic bile ducts	28786	22.07	28938	22.22	0.99	0.99	13580	7.09	12949	5.39	12949	5.39
Bronchus and lung	56851	39.54	56205	39.98	1.01	0.99	24121	13.66	23482	7.86	23482	7.86
Breast	-	-	-	-	-	-	46940.23	44.44	44125	8.68	44125	8.68
Prostate	37905	25.36	36566	24.20	1.04	1.05	-	-	-	-	-	-
Group2: Sites with Age-standardized incidence rate medium												
Lip, oral cavity and pharynx	7579	6.37	7658	6.63	0.99	0.96	2834	1.96	2636	1.91	2636	1.91
Gallbladder	9030	6.16	8972	6.10	1.01	1.01	10454	4.67	10240	4.67	10240	4.67
Pancreas	13209	9.73	13118	9.83	1.01	0.99	11191	5.73	11108	5.69	11108	5.69
Cervix uteri	-	-	-	-	-	-	8961	9.09	8212	10.08	8212	10.08
Corpus uteri	-	-	-	-	-	-	7524	6.64	7179	7.04	7179	7.04
Ovary	-	-	-	-	-	-	8260	7.57	8044	8.14	8044	8.14
Kidney	9251	7.37	9068	7.56	1.02	0.97	4555	2.86	4515	2.92	4515	2.92
Bladder	12542	9.02	12206	8.80	1.03	1.02	3952	1.95	3798	1.87	3798	1.87
non-Hodgkin's disease	7987	6.66	7832	6.85	1.02	0.97	6706	4.54	6550	4.46	6550	4.46
Group3: Sites with Age-standardized incidence rate small												
Larynx	3516	2.69	3368	2.61	1.04	1.03	265	0.16	241	0.18	241	0.18
Skin	556	0.48	537	0.50	1.04	0.97	609	0.45	592	0.47	592	0.47
Testis	1179	1.69	1112	2.21	1.06	0.77	-	-	-	-	-	-
Brain	2257	2.51	2243	2.52	1.01	1.00	2102	1.91	2030	1.83	2030	1.83
Thyroid	1840	1.75	2549	3.57	0.72	0.49	6275	5.94	5649	6.27	5649	6.27
Hodgkin's disease	445	0.51	448	0.56	0.99	0.91	313	0.34	276	0.34	276	0.34
Multiple myeloma	2495	1.75	2486	1.79	1.00	0.98	2357	1.24	2335	1.22	2335	1.22
Leukaemia	5321	5.28	5347	5.34	1.00	0.99	3833	3.54	3725	3.28	3725	3.28
Overall	372752	282.01	362886	282.29	1.03	1.00	275672	197.81	262907	197.02	262907	197.02

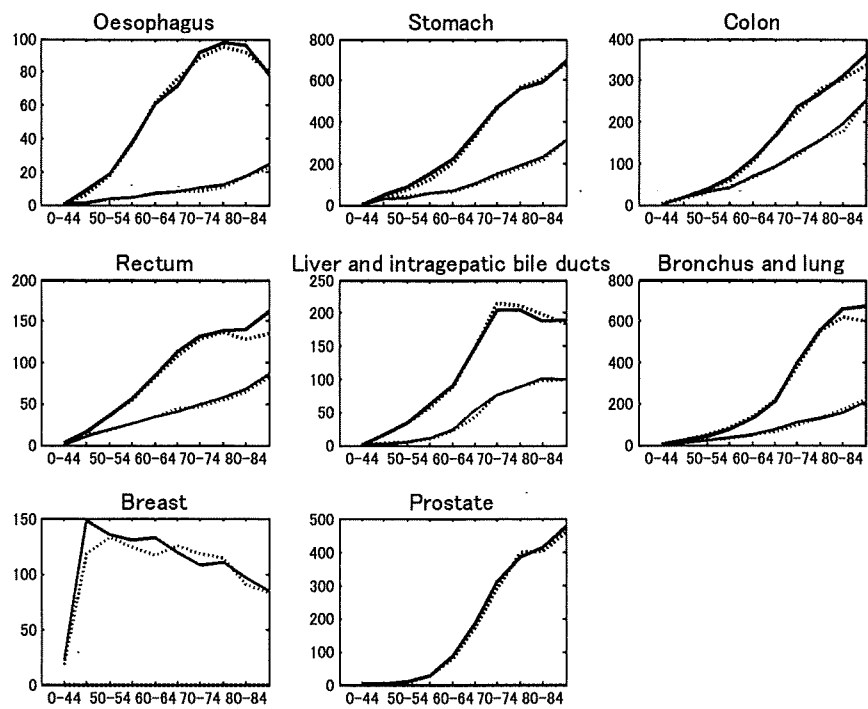


図 1-1. 年齢階級別推計罹患率 (Group 1: 年齢調整死亡率の多い部位)
 実線: 日本法、点線: IARC 法、黒色: 男性、灰色: 女性

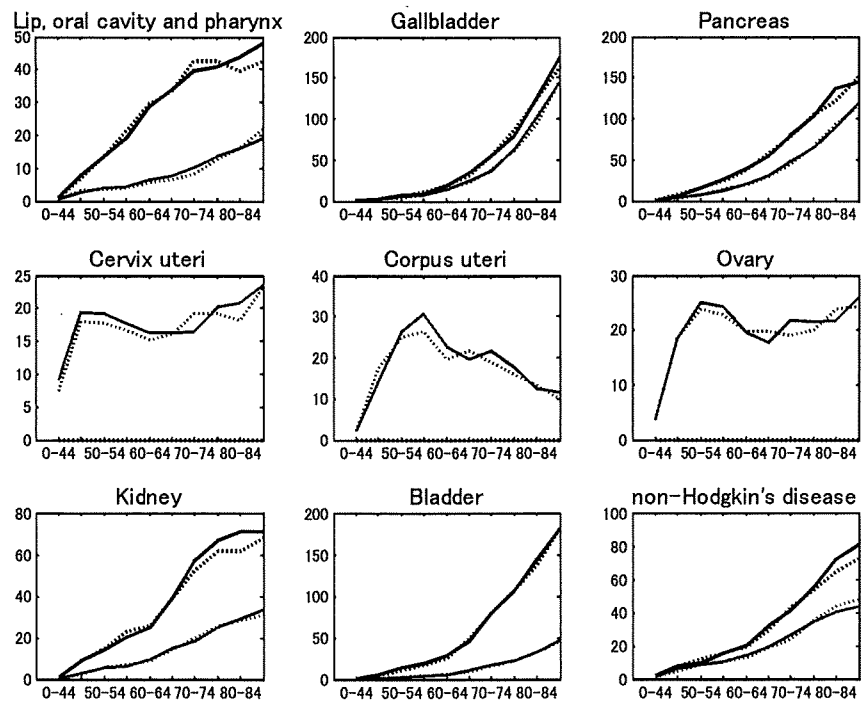


図 1-2. 年齢階級別推計罹患率 (Group 2: 年齢調整死亡率の中程度の部位)
 実線: 日本法、点線: IARC 法、黒色: 男性、灰色: 女性

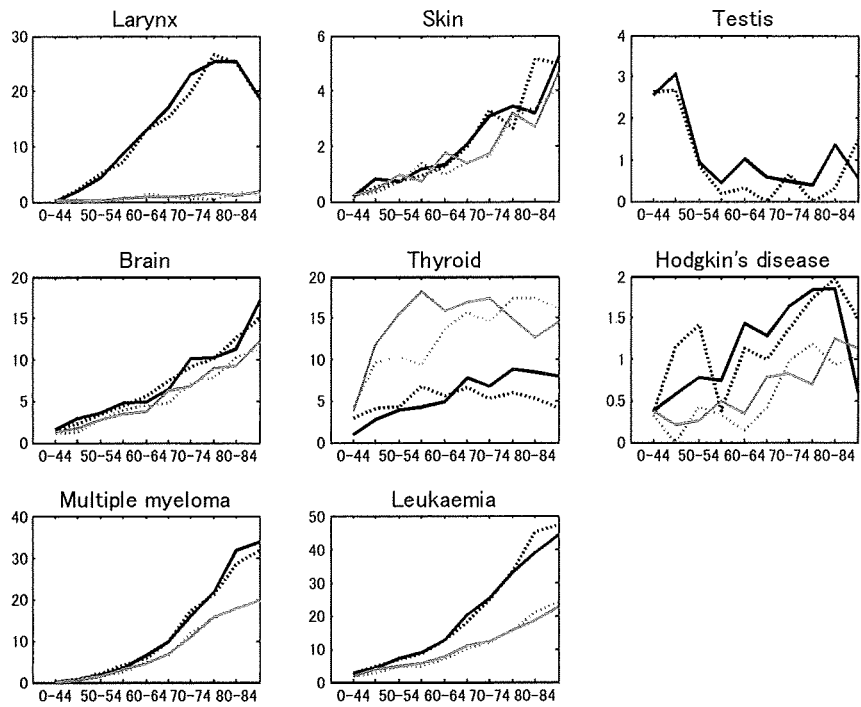


図 1-3. 年齢階級別推計罹患率 (Group3: 年齢調整死亡率の少ない部位)
 実線: 日本法、点線: IARC 法、黒色: 男性、灰色: 女性

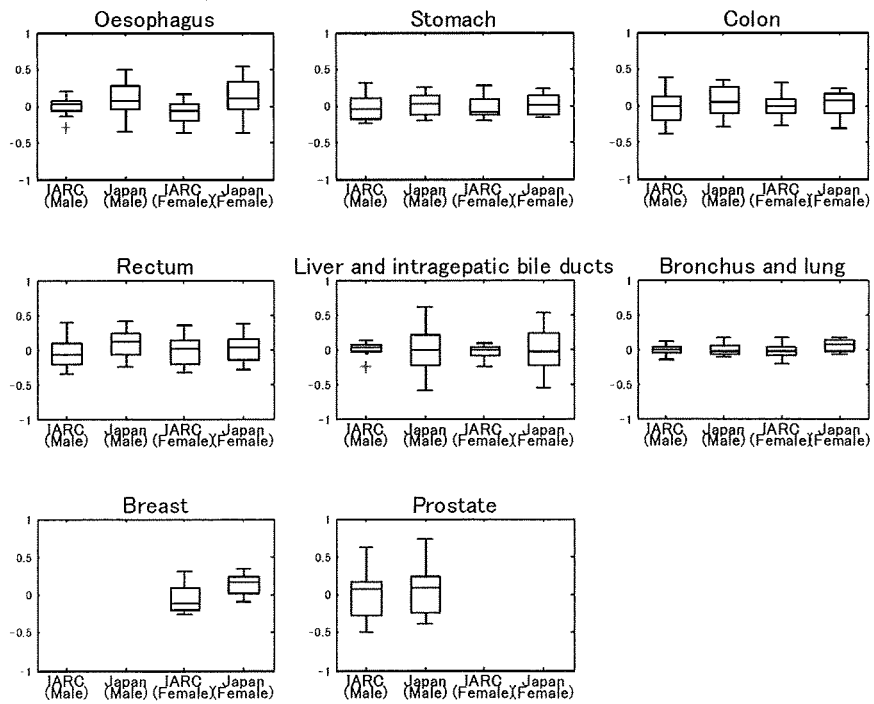


図 2-1. 観測罹患率と推計罹患率の誤差 (Group1: 年齢調整死亡率の多い部位)
 左から、IARC 法 (男性)、日本法 (男性)、IARC 法 (女性)、日本法 (女性)