

患者を抽出すること、また、既登録患者を削除するなどの処理をして登録候補対象患者をできるだけ絞って抽出する作業を行うことが業務となる。

大阪大学医学部附属病院では、がん登録室が医療情報部にあり、がん登録を専門とする診療情報管理士が、ここで集中的にがん登録を行っている。従って、がん登録作業そのものも医療情報部の仕事に含まれている。がん登録を専門とする職員が、どの部署に所属しているのかは病院により様々である。紙の診療録の時代では、コンピュータシステムを管理する医療情報部門と、診療録を扱う病歴管理部門とは異なる組織とされるのが一般的であった。しかし、電子カルテシステムの時代となると、この2つは一体となる必然性が生じる。拡大した医療情報部門では、病院情報システムの企画、運営管理に加え、臨床研究や診療機能評価のためにデータを抽出して提示する業務が加わる。更に、がん登録業務を代表するように、疫学的な調査を目的とした疾患レジストリの登録業務も、医療情報部の業務に加えるのも合理的な形であると考えられる。

E. 結論

がん医療にとって、がん登録は重要である。医療情報部では、がん登録のためのシステムを導入することに加え、月次でがん登録の候補患者をデータウェアハウスから抽出する業務を担う。更に、大阪大学附属病院では、がん登録自身も医療情報部の業務として行っている。

電子カルテの時代となり、病院情報システムを臨床研究でも活用することを推進するべきであり、その推進役として医療情報部が担うべき役割は重要である。登録データが臨床研究に利用できるようにするための仕組み作り、データの抽出業務、更には、疾患レジストリが医療情報部に求められる業務となる。

F. 研究発表

1. 論文発表

1. Chen Y, Matsumura Y, Nakagawa K, Ji S, Nakano H, Teratani T,

Zhang Q, Mineno T, Takeda H. (2007). Analysis of Yearly Variations in Drug Expenditure for One Patient using Data Warehouse in a Hospital, *J Med Syst* 31, 17-24.

2. Matsumura Y, Kuwata S, Yamamoto Y, Izumi K, Okada Y, Hazumi M, Yoshimoto S, Mineno T, Nagahama M, Fujii A, Takeda H. (2007). Template-based data entry for general description in medical records and data transfer to data warehouse for analysis, *Stud Health Technol Inform*, 129, 412-416.
3. Zhang Q, Matsumura Y, Teratani T, Yoshimoto S, Mineno T, Nakagawa K, Nagahama M, Kuwata S, Takeda H. (2007). The Application of An Institutional Clinical Data Warehouse to the Assessment of Adverse Drug Reactions (ADRs): Evaluation of Aminoglycoside and Cephalosporin Associated Nephrotoxicity, *Methods of Inf Med*, 46, 516-522.

2. 学会発表

1. 電子カルテシステムによるがん登録の構築と運用について. 平成 19 年度 大学病院情報マネジメント部門連絡会議抄録集 P138 平成 20 年発行
2. 大阪大学医学部附属病院における DWH のデータ抽出業務の依頼事例. 診療録管理第二十巻第二号 P188 平成 20 年発行
3. 臨床研究の支援を目指したデータウェアハウスの構築. 第 28 回医療情報学連合大会論文集 平成 20 年発行

複数データベースの結合における問題点について

—多施設共同研究における情報伝達・収集・管理上の問題点・留意点—
—異種データベース連結における情報管理・連結上の問題点—

研究協力者 沼崎 穂高（大阪大学大学院医学系研究科）

研究要旨

現在日本で運用されている主ながん登録である「地域がん登録」、「臓器別がん登録」、「院内がん登録」に関係するいくつかのデータベースのデータフォーマットやデータ登録システムを調査し、他施設からのデータ収集や、それぞれのデータベース間のデータ連携を行う上での問題点を抽出し、その対応策を示唆した。

A. 研究目的

他施設共同研究や全国集計などを行う際に発生するデータ連携、管理の問題点を検討する。

B. 研究方法

以下に挙げるがん診療データを扱うデータベースのデータフォーマットやデータ登録システムを調査し、他施設からのデータ収集や、それぞれのデータベース間のデータ連携を行う上での問題点を抽出する。

- (1) 日本放射線腫瘍学会 (Japanese Society of Therapeutic Radiology and Oncology: JASTRO)のデータベース委員会が放射線治療施設に対して提示している放射線治療情報の標準フォーマットを用いた放射線腫瘍学広域データベース (Radiation Oncology Greater Area Database: ROGAD)。著者は同委員。
- (2) 厚生労働省科学研究費補助金第3次対がん総合戦略事業手島班で構築、運用している Japanese National

Cancer Database (JNCDB)のフォーマットと登録データベースソフト。著者は同班の分担研究者。

- (3) 日本食道学会 (The Japan Esophageal Society: JES)の全国登録委員会が実施している食道癌全国登録のフォーマットと登録データベースソフト。著者は同委員。
- (4) 厚生労働省が推進しているがん診療連携拠点病院の院内がん登録のフォーマットと登録データベースソフト。

(倫理面への配慮)

ROGAD, JNCDB, 院内がん登録に関して、本研究では各データベースのデータフォーマットと登録データベースソフト自体を調査するため、実際に患者のデータを取り扱うことはない。従って倫理上の問題は発生しない。

食道癌全国登録に関して、著者が開発した個人情報匿名化ソフトを用いて個人情報部分を「連結不可能匿名化」したデータを各施設からCD-Rで収集するため「疫学研究に関する倫理指針」の適用外となり、個人情報も取り扱わない。よつ

て患者のプライバシーの確保に関しても問題ないと考えられるが、症例データの管理に関して個人情報と同等の安全性と守秘性を確保する。

C. 研究結果

1. データ連携のための一部フォーマットの統一化

ROGAD, JNCDB, 食道癌全国登録データベースはそれぞれ各施設の診療科データベースとしての機能を有しており、院内でのデータ連携を図るために全データベースの共通部分である患者個人情報のデータフォーマットを院内がん登録の標準登録様式と統一させている。

また、放射線治療情報を基盤とする JNCDB は疾患共通部分である基本データベースと疾患固有部分である各論データベースに分かれており、疾患共通部分である基本データベースはデータ連携を容易にするために ROGAD と同じデータフォーマットを採用している。さらに疾患固有部分の内、食道癌データベースは食道癌全国登録のデータフォーマットを網羅した形式としている。

2. 全国集計（他施設共同研究）におけるデータ収集の問題点の抽出

食道癌全国登録では 2001 年の食道癌治療例を全国 241 施設から 4458 症例収集した。データ収集、データクリーニングで発生した主な問題点を以下に記載する。

- a) データ不備
収集データの文字化け
収集データファイルの破損 等
- b) 登録年の不整合
治療開始日が 2001 年ではないが登録年が 2001 年である症例
- c) 日付の不一致
最終生存確認日 < 治療開始日
診断日 > 治療開始日 等
- d) 必須項目の未入力
生年月日、性別 等

3. 異種データベース間のデータ連携の問題点の抽出

各データベースのフォーマットと使用方法を調査し、それぞれのデータベース間でデータ連携を行う際の主な問題点を以下に記載する。

a) データの信憑性

異なるデータベースに同一症例が格納されていた場合、一方のデータベースにないデータは保管することは可能であるが、両データベースの同一項目が異なる値で入力されている場合、どちらのデータを真とするかが問題となる。

b) 選択肢のデータの整合性

異なるデータベース間では項目が同一でも選択項目が異なる場合があり、その際のデータの整合性が問題となる。

D. 考察

現在、日本で運用されているがん診療に関するデータベース、いわゆるがん登録は大きく以下の 3 つに分けられる。自治体主導の「地域がん登録」、学会主導の「臓器別がん登録」、施設主導の「院内がん登録」である。この内、「院内がん登録」と「地域がん登録」は国立がんセンターのがん対策情報センターが中心となってデータフォーマットの統一や登録の方法、腫瘍登録士の育成などを推進しており、多施設や多地域の登録データの収集、管理、解析の運用をしている。

一方で今回調査対象とした JNCDB や食道癌全国登録などの臓器別がん登録は学会や、研究班が母体となり進めているものである。ROGAD, JNCDB と食道癌全国登録のようにそれぞれの学会、研究班が協力してデータフォーマットの整合性やデータ連携のための共同研究を行っている場合もあるが、多くは臓器毎、つまり学会毎にデータフォーマットの取り決めやデータベースシステムの構築を行っており、他のデータベースとのデータ連携は思うように進められていない。

食道癌全国登録の調査結果からも判明

したようにひとつのデータフォーマットとデータベースソフトを用いて全国集計を行う場合でも、各施設間でデータの入力方法が異なったり、施設間のデータの正確さが様々であったりすることにより、収集したデータが解析対象とならなかったり、欠損データが多くなることで解析結果に影響を及ぼす可能性が生じることがある。この問題を打開するためには登録を行う施設へのデータフォーマットや登録方法に関する啓蒙活動が必要であり、また各施設でのデータ入力担当者（現在は多くの場合で臨床医）の入力負担を軽減し、入力補助を行える機能を付加した診療科データベースの構築が必須となる。

多くの施設の診療科データベースは臨床医が忙しい臨床業務の合間に構築した各科独自のものや学会が準備した登録ソフトとなっており、電子カルテや院内がん登録システムのように施設が管理しているものではない。一方でがん患者が欲している最も正確で詳細な診療データが入力されているのは、現時点では、施設が管理する電子カルテや院内がん登録システムではなく、各科が独自に管理する診療科データベースである。同一施設内の電子カルテ、院内がん登録、各診療科データベースのデータ連携がスムーズに流れれば、各施設から出てくるデータの正確性も担保され、それぞれの全国集計結果の信憑性も向上することが期待できる。そのためには現在各科が独自で運用している診療科データベースを施設の中核で管理でき、さらに各診療科の診療や研究活動に支障をきたさずにデータ提供できる仕組みが必要となってくる。また各データベース間でデータフォーマットの対応表を作成する等の取り組みも同時に必要である。

上記のように各施設内でのデータ連携がスムーズに進めば、異種データベースの全国集計間でのデータ連携の問題点であるデータの信憑性、選択肢のデータの整合性の問題も解決されることとなり、がん患者が真に求めるデータを全国的なデータとして提示することが可能となる。

E. 結論

がん診療データを扱う様々なデータベースのデータフォーマットやデータ登録システムを調査し、他施設からのデータ収集や、それぞれのデータベース間のデータ連携を行う上での問題点を抽出し、その対応策を示唆した。

F. 研究発表

1. 論文発表

Uno, T., Sumi, M., Ishihara, Y., Numasaki, H., Mitsumori, M., Teshima, T., Japanese PCS Working Subgroup of Lung Cancer. (2008). Changes in patterns of care for limited-stage small-cell lung cancer: Results of the 99-01 patterns of care study - A nationwide survey in Japan, *International Journal of Radiation Oncology, Biology, Physics*, 71(2), 414-419.

Teshima, T., Numasaki, H., Shibuya, H., Nishio, M., Ikeda, H., Ito, H., Sekiguchi, K., Kamikonya, N., Koizumi, M., Tago, M., Nagata, Y., Masaki, H., Nishimura, T., Yamada, S., Japanese Society of Therapeutic Radiology and Oncology Database Committee. (2008). Japanese structure survey of radiation oncology in 2005 based on institutional stratification of patterns of care study. *International Journal of Radiation Oncology, Biology, Physics*, 72(2), 144-152.

Numasaki, H., Ohno, Y., Ishii, A., Kasahara, S., Fujimoto, H., Harauchi, H., Inamura, K., Monden, M., Sakon, M. (2008). Workflow Analysis of Medical staffs in Surgical Wards Based on the Time-Motion Study Data. *Japan Hospitals*, 27, 75-80.

Numasaki, H., Teshima, T., Shibuya, H., Nishio, M., Ikeda, H., Ito, H., Sekiguchi, K., Kamikonya, N., Koizumi, M., Tago, M., Nagata, Y., Masaki, H., Nishimura, T., Yamada, S., Japanese Society of Therapeutic Radiology and Oncology Database Committee. (2009). National Structure of Radiation Oncology in Japan with Special Reference to Designated Cancer Care Hospital. International Journal of Clinical Oncology. (*in press*).

小澤 壯治, 日月 裕司, 田中 乙雄, 篠田 雅幸, 宇田川 晴司, 松原 久裕, 馬場 秀夫, 竹内 裕也, 小山 恒男, 室 圭, 宇野 隆, 手島 昭樹, 沼崎 穂高, 山名 秀明, 小西 敏郎. (2008). 食道癌全国登録の再開にあたり－問題点と解決法－. 癌と化学療法, 35(9), 1497-1499.

2. 学会発表
該当なし

G. 知的所有権の取得状況

1. 特許取得
該当なし
2. 実用新案登録
該当なし
3. その他
該当なし

H. 文献

日本放射線腫瘍学会ホームページ
ROGAD 掲載ページ
(<http://www.jastro.jp/report/topic/070423.html>)

日本食道学会ホームページ 食道癌全国登録掲載ページ
(http://www.jncdb.org/jes/Esophagus_Top.html)

国立がんセンター がん対策情報センターホームページ がん情報サービス 院内がん登録掲載ページ
(http://ganjoho.ncc.go.jp/hospital/cancer_registration/index.html)

患者データのセキュリティにおける質保証

—IAEAにおけるデータベース情報のセキュリティ管理について—

研究協力者 沼崎 穂高（大阪大学大学院医学系研究科）

研究要旨

患者データを含む情報を管理する上では強固なセキュリティ管理が必要であり、また国際的なデータを公表するためにはデータの精度管理が重要となってくる。そのためには人的資源と費用の必要性、また他国のデータを統括して管理する上でのデータの標準化が重要となってくる。

A. 研究目的

国際原子力機関（International Atomic Energy Agency: IAEA）でのデータ管理業務を振り返り、データ管理をする上で重要となるセキュリティ管理と精度管理について考察する。

B. 研究方法

IAEAでは144の加盟国の放射線治療機器保有施設の構造情報を集積し公開している（Directory of Radiotherapy Centres: DIRAC）。日本では日本放射線腫瘍学会（Japanese Society for Therapeutic and Oncology: JASTRO）が学会事業として全国放射線治療施設構造調査を行っており、2005年の定期調査からDIRACへのデータ提供を行っている。IAEAからJASTROへの日本のデータの管理を行う人材の派遣要請を受け、定期構造調査のデータを管理・解析していた筆者が2008年8月～10月の3ヶ月間IAEAに在籍し、DIRACの日本のデータの整理と更新の業務に携わった。

IAEAでの業務で得られたデータのセキュリティ管理と精度管理の重要性を考察する。

（倫理面への配慮）

特になし

C. 研究結果

契約開始時点でDIRAC内には920の

日本の放射線治療機器保有施設が登録されていた。3ヶ月間での業務は（1）DIRAC内データの整理、（2）整理されたデータの更新と新規施設データの登録の2つに分けられた。結果として契約終了時にはDIRAC内の放射線治療機器保有施設は791施設となった。

D. 考察

DIRACの全データは1人のデータベースマネージャが管理している。セキュリティ管理上はリスク軽減となるが、定期的なデータ更新をしておらず、施設名の変更や住所変更に伴う2重登録や放射線治療機器が既に廃棄されている施設の登録がある等、データの精度が低かった。

国際的なデータを取り扱う以上、精度向上が必須であるため、改善の為に今回のように定期的に加盟各国からデータ管理業務を行う人材を短期で雇用しており、多くの人的資源と国際的なデータ管理業務に見合う給与が支給されている。

一方で、筆者ら短期雇用のデータマネージャが実業務で扱うデータは自国のデータのみとなり、そのデータのDIRACへの登録は上述のデータベースマネージャが1で行うなど中枢のデータのセキュリティ管理は徹底されている。また当然のことながら、公表されたデータ以外の業務で得られたデータは口外しないことが契約書に記載されている。

DIRAC のデータフォーマットは放射線治療機器の製造会社や型番や施設の放射線治療部門のスタッフ数、年間患者数など、非常に詳細で多岐に渡っているが、実際に詳細なデータを登録している国は日本を含めた数カ国であり、他国のデータは欠損が多いため、全体集計では治療機器台数、スタッフ数、患者数など数項目のみしかない。より詳細な全体集計結果を公表するためにはデータのみならず、登録方法自体の標準化が必須である。

E. 結論

国際機関でのデータ管理業務の経験を基にデータ管理をする上で重要となるセキュリティ管理と精度管理の重要性を示唆できた。

F. 研究発表

1. 論文発表

Uno, T., Sumi, M., Ishihara, Y., Numasaki, H., Mitsumori, M., Teshima, T., Japanese PCS Working Subgroup of Lung Cancer. (2008). Changes in patterns of care for limited-stage small-cell lung cancer: Results of the 99-01 patterns of care study - A nationwide survey in Japan, *International Journal of Radiation Oncology, Biology, Physics*, 71(2), 414-419.

Teshima, T., Numasaki, H., Shibuya, H., Nishio, M., Ikeda, H., Ito, H., Sekiguchi, K., Kamikonya, N., Koizumi, M., Tago, M., Nagata, Y., Masaki, H., Nishimura, T., Yamada, S., Japanese Society of Therapeutic Radiology and Oncology Database Committee. (2008). Japanese structure survey of radiation oncology in 2005 based on institutional stratification of patterns of care study. *International Journal of Radiation Oncology, Biology, Physics*, 72(2), 144-152.

Numasaki, H., Ohno, Y., Ishii, A., Kasahara, S., Fujimoto, H., Harauchi, H., Inamura, K., Monden, M., Sakon, M. (2008). Workflow Analysis of Medical staffs in Surgical Wards Based on the Time-Motion Study Data. *Japan Hospitals*, 27, 75-80.

Numasaki, H., Teshima, T., Shibuya, H., Nishio, M., Ikeda, H., Ito, H., Sekiguchi, K., Kamikonya, N., Koizumi, M., Tago, M., Nagata, Y., Masaki, H., Nishimura, T., Yamada, S., Japanese Society of Therapeutic Radiology and Oncology Database Committee. (2009). National Structure of Radiation Oncology in Japan with Special Reference to Designated Cancer Care Hospital. *International Journal of Clinical Oncology*. (in press).

小澤 壯治, 日月 裕司, 田中 乙雄, 篠田 雅幸, 宇田川 晴司, 松原 久裕, 馬場 秀夫, 竹内 裕也, 小山 恒男, 室 圭, 宇野 隆, 手島 昭樹, 沼崎 穂高, 山名 秀明, 小西 敏郎. (2008). 食道癌全国登録の再開にあたり - 問題点と解決法 -. *癌と化学療法*, 35(9), 1497-1499.

2. 学会発表

該当なし

G. 知的所有権の取得状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし

H. 文献

該当なし

がん登録の悉皆性評価に関するサーベイ

研究協力者 志岐 直美（大阪大学大学院医学系研究科）
研究代表者 大野ゆう子（大阪大学大学院医学系研究科）

研究要旨

地域がん登録は国レベルでのがん罹患の動向を把握できる唯一のシステムである。諸外国ではがん登録の実施が法で義務付けられるなどして、高い登録率を維持し、医療政策に活用されている。我が国においても同様の活用が期待されるが、我が国におけるがん登録は未だ発展途上の段階にある。がん登録が確立していない現状においては、がん登録の悉皆性の評価、および欠損値の補完が重要課題である。そこで、本報告では地域がん登録の悉皆性評価について概説する。

A. 研究目的

がん登録とはがん患者の診断、治療、予後に関する情報を収集する仕組みであり、がん対策の立案、評価に必要な情報を提供する。がん登録は地域がん登録、院内がん登録、臓器がん登録の3種類に分けられる。なかでも地域がん登録は国レベルでのがん罹患の動向を把握できる唯一のシステムである。諸外国におけるがん登録は、その実施が法で義務付けられるなどして高い登録率が維持され、医療政策に活用されている。

一方、我が国では2007年に閣議決定された「がん対策推進基本計画」において、医療機関等の整備を図るなかで重点的に取り組むべき項目の一つとして「がん登録」の推進が掲げられた。これにより初めて、地域がん登録および院内がん登録は、実質的なサーベイランスを担う仕組みとして、行政として積極的に取り組むべき課題であると解釈されることとなった。

しかし2008年現在、東京都をはじめいくつかの自治体は地域がん登録を実施しておらず、実施報告は35道府県と1市（広島市）にとどまっている。しかも、その登録精度は国際的な基準からすれば不十分な地域が多く、国際がん研究機関（International Agency for Research on Cancer, IARC）が5年に1回刊行している「5大陸のがん罹患（Cancer Incidence in Five Continents）」の最新巻（第9巻）

では、我が国からは6府県と1市（宮城県、山形県、福井県、愛知県、大阪府、広島市、長崎県）の地域がん登録のみが、比較的登録精度が高い地域がん登録として収録されている（Curado et al.(2007)）。国際がん登録学会において、日本はがん登録の開発途上国とされている。

がん医療の実態把握、がん対策の立案、評価のために、がん登録資料の活用が期待されるが、がん登録が確立していない我が国の現状においては、がん登録の悉皆性（完全性）の評価、および欠測値の補完が重要課題である。

そこで本報告では、がん登録の悉皆性（完全性）の評価方法および登録率の推計方法、欠損値の補完方法について概説する。

B. 研究方法

既存資料を基に、1.我が国における地域がん登録の精度指標、2.登録率の推計方法、3.欠損値の取り扱い方法、について情報収集しまとめた。

C. 結果および考察

1. 我が国における地域がん登録の精度指標

地域がん登録の精度指標に関する概念図を図1に示す。地域がん登録において、がん罹患の届出患者ファイルとがん死亡者ファイルとを照合した時点で、罹患の届出がなくがん死亡票によって初めて把

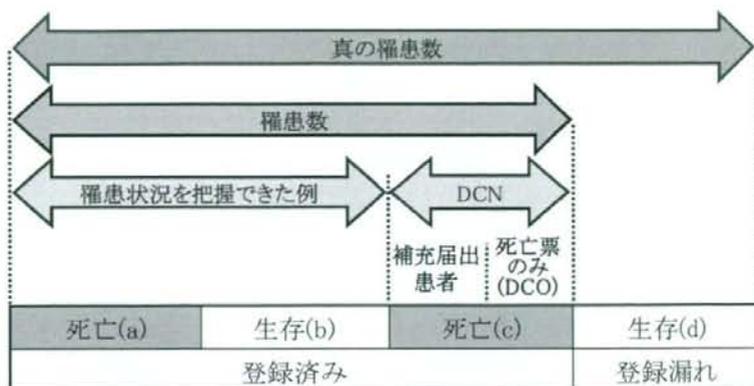


図1 地域がん登録の精度指標

祖父江 他(2007)より一部改変

死亡数=死亡(a)+死亡(c)

罹患数=死亡(a)+生存(b)+死亡(c)

DCN:Death Certification Notification、死亡票により初めて把握された患者

DCN%=死亡(c)/(死亡(a)+生存(b)+死亡(c))

DCO:Death Certification Only、死亡情報のみで登録された患者

DCO%=死亡(c)内のDCO/(死亡(a)+生存(b)+死亡(c))

IM比=(死亡(a)+生存(b)+死亡(c))/(死亡(a)+死亡(c))

握された者は Death Certification Notification (DCN)と呼ばれる。DCN 症例については生前の医療情報の収集を試み、情報が得られたものは補充届出患者として登録される。一方、生前情報が得られなかった症例は Death Certificate Only (DCO)として、死亡日を罹患日とみなし、死亡情報のみで登録される。

地域がん登録の精度における悉皆性の指標として、「DCN 割合(DCN%)」、「DCO 割合(DCO%)」、「Incidence Mortality Ratio: IM 比(罹患死亡比)」がある(祖父江 他(2007))。

DCN%とは、ある地域がん登録室で観察された対象期間中の全罹患患者に占める DCN 症例の割合である。この値が大きいほど、図1中の(d)に示される登録漏れ・把握不可能な生存患者が多く存在することが推測される。この登録漏れ患者については、患者が死亡するまでは地域がん登録で把握できないと推察され、その分地域がん登録の罹患数は真の罹患数より低く見積もられていると考えられる。

DCO%とは、ある地域がん登録室で観察された対象期間中の全罹患患者に占める DCO 症例の割合であり、間接的に悉皆性を反映するものである。DCO%が高けれ

ば登録の悉皆性が低いといえるが、DCO%が低いからといって必ずしも悉皆性を保証するものではない点に留意する必要がある。なぜなら DCN%が高くて生前情報の収集に力を注げば DCO%を低下させることが可能だからである(Brenner(1995))。

また、IM 比は罹患数と死亡数の比である。予後の悪いがんでは1に近い値をとり、予後の良いがんでは1より大きな値をとる。

我が国におけるがん罹患率、罹患数の全国集計を行う際の集計対象の基準として、DCO%が25%以下もしくは DCN%が30%未満、かつ IM 比が1.5以上という基準がある。2002年のがん罹患の報告では、11の地域がん登録(宮城県、山形県、神奈川県、新潟県、福井県、滋賀県、大阪府、鳥取県、岡山県、佐賀県、長崎県)のみがこの基準を満たしており、全国推計に用いられている(Matsuda et al.(2008))。今後、他地域がん登録においても更なる精度向上が望まれる。

2. 登録率の推計方法

地域がん登録における登録率の推計方法には flow method (Bullard et

al.(2000)), capture-recapture method のほか (Schouten et al.(1994)、Schmidtman(2008))、経時データ法として、罹患率の安定性や異なる集団間の罹患率の比較、年齢別罹患率曲線の検討がある (Parkin(2006))。なかでも一般的な登録率の推計方法として、DCN%とIM比を用いた推計式があり、Parkin et al.(1994)、味木 他(1998)、加茂(2008)により3つの登録率の推計式が提案されている。以下簡単に説明する。

我が国では人口動態統計があるため、死亡者については漏れなく把握できていると仮定した場合、真のがん罹患数は以下の4つの患者群から構成されていると考えることができる。即ち、地域がん登録に、登録されかつ死亡している患者数 a 、登録されかつ生存している患者数 b 、登録されておらずかつ死亡した患者数 (DCN) c 、登録されておらずかつ生存している患者数 d (不明)である。そのため、地域がん登録の登録率 r 、DCN% x 、IM比 y は下記で示される。

$$r = \frac{a+b+c}{a+b+c+d},$$

$$x = \frac{c}{a+b+c},$$

$$y = \frac{a+b+c}{a+c}$$

DCN%およびIM比を用いた登録率の推計方法は、初め Parkin により提案された (Parkin et al.(1994))。Parkin の方法では以下の3つの仮定を要する。

- (1) 届出がなかった患者における死亡割合 $c/(c+d)$ は、届出患者における死亡割合 $a/(a+b)$ に等しい。よって、 $d=bc/a$ である。
- (2) その年の罹患患者における死亡数 $a+c$ は、その年のがん死亡数にほぼ等しい。
- (3) DCN%が10%程度と比較的小さい、即ち c が小さい場合、次式が成立する。

$$\frac{a}{a+b} \cong \frac{a+c}{a+b+c} = \frac{1}{y}$$

上記3つの仮定から、Parkin は登録率 r の推計式を以下のように提案した。

$$r = \frac{1}{(1-x) + \frac{x}{1/y}} \quad (1)$$

一方、我が国では欧米に比してDCN%が高い地域がん登録室が多いため、仮定(3)が当てはまらない場合が多い。そこで、味木らは仮定(3)を用いない以下の登録率の推計式を提案した (味木 他(1998))。

$$r = \frac{1-xy}{1-x} \quad (2)$$

なお、Parkin、味木らによる式(1)、(2)の詳しい算出方法は Appendix に示す。

近年では更に、加茂らにより数学モデルを用いた真の罹患数を推定する方法が提案されている (Kamo et al.(2007)、加茂(2008))。本方法はDCN%とIM比の間にある非線形関係を用いた回帰モデルを構築し、真のIM比を推定してそれに死亡数を乗ずることで真の罹患数を推計している。DCN%が0 (即ち、登録率100%) のときのIM比を真のIM比 β (未知)とすると、以下の関係式が成立する。

$$y = \frac{\beta}{1+(\beta-1)x} \quad (3)$$

即ち、真のIM比 β が性別、原発部位別で、全地域がん登録室間で一定と仮定した場合、ある地域がん登録室 i で観察されるIM比 y_i とDCN% x_i は次式の関係をもつ。

$$E(y_i|x_i) = \frac{\beta}{1+(\beta-1)x_i} \quad (4)$$

最尤法を用いて β を推定する。

加茂らはここで、仮定(1)を拡張し、既に登録されている患者群と登録されていない患者群における死亡割合の差異を示す未知パラメータを α (未登録群における死亡割合は既登録群より α 倍高い) とし、 α と β を同時に推定するモデルを構築している。

地域がん登録室で観察されるがん罹患数を n_i 、死亡数を m_i 、IM比の逆数を v_i とすると、 $v_i = m_i/n_i$ である。がん死亡は、がん罹患のうち確率 v_i で発生すると考えられるので、 m_i は2項分布に従い、以下のように表現できる。

$$m_i | x_i \sim \text{Bin}(n_i, v_i) \quad (5)$$

$\alpha=1$ であれば v_i は式(3)で表現されるが、 $\alpha \neq 1$ でない場合、 v_i は次式での2つの解のうち大きい方で表される。

$$\beta v_i^2 + (\alpha x_i - \beta x_i - 1)v_i + (1 - \alpha)x_i = 0 \quad (6)$$

ただし、 $\alpha > 1$ 、 $\beta > 1$ である。

そこで、 $a = \log \alpha$ 、 $b = 1 + \log \beta$ とおき、未知パラメータベクトル $\theta = (a, b)'$ とおくと、対数尤度関数は

$$n \sum w_i \left[v_i \log \left\{ \frac{v_i(\theta)}{i - v_i(\theta)} \right\} + \log(1 - v_i(\theta)) \right] \quad (7)$$

となる。ここで、 n は対象地域における全罹患数 $\sum n_i$ であり、 $w_i = n_i / n$ である。つまり、未知パラメータ推定における地域ウエイトは、各地域が登録室で確認された罹患数が全罹患数に占める割合に基づいて設定される。罹患数は、地域規模にある程度比例し、また登録の悉皆性が高ければ増える数であるので、これら2つを同時に満たすウエイトと考えることができる。

$\alpha=1$ 、もしくは $\alpha \neq 1$ いずれのモデルを選択するかは、交差検証法 (Cross-validation) を用いている。具体的には、以下の値を各モデルにおいて算出し、小さい値をとるモデルを採択する。

$$CV = -2 \sum \left\{ m_i \log p_{i-1} + (n_i - m_i) \log(1 - p_{i-1}) \right\} \quad (8)$$

ここで、 p_{i-1} は、 v_i のジャックナイフ推定量を表し、第 i 番目の地域がん登録室の観測値を除いて推定した値である。

本章では3種類、Parkin、味木、加茂の登録率の推計式を紹介したが、登録率は Parkin の式で最も過大評価され、味木、加茂の式と続く。これらの方法を用いた罹患数の補正值の公表については未だ議論の余地があるとされている。特に、上記3つの登録率の推計式では年齢構成が考慮されていない。IM比は年齢の影響を大きく受けると推察され、年齢階級によって登録率、DCN%が異なることが予想される。年齢階級別の登録率推計など、年齢の影響を考慮したモデルの構築が望まれる。

3. 欠損値の取り扱い方法

欠測値は、欠測の発生する理由により大きく以下の3つに分類される。1つは Missing Completely at Random

(MCAR) と呼ばれるもので、欠測するかどうかは完全にランダムであり、予測不可能な欠測を指す。1つは Missing at Random (MAR) と呼ばれるもので、欠測する変数 Y の欠測のメカニズムが前回の観測値や他の変数 X には依存し得るが変数 Y 自身には依存しない欠測を指す。最後の1つは Missing Not at Random (MNAR) と呼ばれるものであり、欠測はランダムではなく、欠測するか否かが欠測する変数 Y そのものに依存しており、他の変数 X から予測不可能なもの指す。欠測値の種類に応じて適切な解析手法を選択する必要がある。

欠測値を含むデータの解析方法として、まず、欠測値を含むデータを除いて分析する完全事例解析 (complete case analysis) がある。この方法は欠測値の発生が全体の5%以下と少ない場合には、簡便な解析方法として利用可能である。ただし、本解析方法は欠測値の存在によるサンプリングバイアスについては調整できない。

また、欠測のない完全なデータのみでなく、得られたデータを全て用いて分析する手元のデータによる解析 (available case analysis) がある。この方法では、集約した統計量が偏りをもつ可能性や、得られる分散共分散行列や相関行列が正定値でなくなり、多変量解析に適さなくなる可能性があるという問題がある。

そこで欠測値に何らかの値を補完して解析する方法が代入法 (imputation) である。代入法には欠測箇所の一つの値を埋め込む 1) 単一値代入法 (single imputation) と、複数個の値を代入し擬似的な完全データを作成する 2) 多重代入法 (multiple imputation) がある。単一値代入法には、(1) 平均値代入法、(2) Indicator method、(3) 重回帰式、(4) Hotdeck 法がある。以下、簡単に説明する。

1) 単一値代入法

- (1) 平均値代入法：欠測値に、観察値の平均値を代入する方法。この方法では真のデータより標準誤差が小さくなる。
- (2) Indicator method：回帰分析を行う前に、欠測値をもつカテゴリ i ごとに $K_i + \beta_j X_i (1 - K_i)$ で示される指標

をつくり、変数 X_i が欠測しているとき $K_i = 1$ 、 X_i が欠測していなければ $K_i = 0$ とする。即ち、欠測値を示すカテゴリを作成する方法。

- (3) 回帰分析による代入法：欠測値を持たないケースのデータを用いて回帰分析を行い、予測値を欠測値に代入する方法。例えば表1のデータセットがある場合、説明変数を性別(Sex: 1 男性、2 女性)、目的変数を年齢(Age)として重回帰分析を行うと、表2の結果が得られる。得られたモデルから欠測値を予測することで、表1の補完値が得られる(Nur(2008))。
- (4) Hotdeck法：欠測値をもつケースと似た背景データをもつケースを特定し、そのケースの値を欠測値に代入する方法。問題点として、ケースが「似ている」という定義が難しいこと、欠測値の特性が不明なままであることなどが挙げられる。例として表3では、年齢

に欠測値をもつケースが確認されたが、性別、進行度、部位が同じケースの年齢を欠測値に補完している。

2) 多重代入法

多重代入法では、欠測値が MAR であると仮定する。単一値代入法では、推定量のばらつきを過小評価するという問題がある。これを克服するために、多重代入法では複数回の補完を行うことによりこの問題解決を図る。手順としては3段階あり、(1)欠測値に M 個の異なる値を代入し、M 個の擬似完全データセットを生成する。(2) M 個の擬似完全データセットそれぞれに、完全データセット用の解析手法を適用して分析する。(3) M 種類の分析結果を一つに統合する、の順に解析を進める。多重代入法としては、Buuren et al.(1999) が報告した Multivariate Imputation by Chained Equation (MICE)法がある。MICE法では、まず元データセットのコピーを3から5つ作成し、それぞれについて欠測値

表1 仮想データ:補完前後の値

性別	年齢	補完値
男	.	71.95
女	.	71.58
女	.	71.58
女	67.30	67.30
女	72.26	72.26
男	58.98	58.98
女	75.18	75.18
男	83.63	83.63
男	78.16	78.16
男	67.01	67.01

Nur(2008)より抜粋。

表2 重回帰モデル

$$Age = \alpha + \beta Sex + \varepsilon$$

変数	係数	標準誤差	p 値
切片	72.31	10.32	<0.01
性別	-0.37	6.83	0.96

Nur(2008)より抜粋。

表3 仮想データ:Hotdeck法による欠測値補完例

ID	性別	進行度	年齢	部位
1	男	所属リンパ節転移	27	胃
2	男	隣接臓器浸潤	56	肺
3	女	限局	33	乳房
4	女	隣接臓器浸潤	55	大腸
5	男	遠隔転移	65	肺
6	女	隣接臓器浸潤	55	大腸
7	男	所属リンパ節転移	27	胃
8	男	遠隔転移	50	肝臓
9	女	限局	40	胃
10	男	隣接臓器浸潤	56	肺

灰色のセルが欠測値であった部分。性別、進行度、部位で観測値が同じケースがあれば、そのケースの値を補完する。

表3では、ID=7の年齢にはID=1の年齢を、ID=10の年齢にはID=2の年齢が補完されている。

を適当な方法により補う。これらの擬似完全データセットに検討したいモデルをあてはめ、それぞれからパラメータの推定値を得る。得られたパラメータの推定値の平均を元データセットのパラメータ推定値と見做し、その標準誤差を求める。 M 個の擬似完全データセットから推測対象であるパラメータ θ の推定値 $\hat{\theta}_1, \dots, \hat{\theta}_M$ 、およびそれらの標本分散 U_1, \dots, U_M が得られたとする。このとき、 θ の点推定値は算術平均

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \theta_m$$

で与えられる。このとき、標本分散 T は代入内分散

$$\bar{U} = \frac{1}{M} \sum_{m=1}^M U_m$$

および代入間分散

$$B = \frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}_m - \bar{\theta})^2$$

を用いて $T = \bar{U} + \left(1 + \frac{1}{M}\right) B$ と計算される。

例として、表4のようにID、胸囲、肺活量に関する仮想完全データを設定する。更に、胸囲が80cm以上で肺活量が欠測すると仮定した欠測データを作成する。Markov Chain Monte Carlo (MCMC)法を用いて欠測値を補完した5つの擬似完全データセットを作成したのち、説明変

数を胸囲、目的変数を肺活量とした重回帰分析を行った場合、表5のモデル1から5までの結果が得られる。このモデルにより得られた推定値の平均および分散を統合して得られたモデルが表5内の統合モデルの結果であり、そのモデルから欠測値を補完した値が表4内の補完値である。

登録精度が不十分な地域がん登録データにおいては、年齢や進行度が不明の場合が多い。しかし、欠測値のないデータのみを用いた解析(完全事例解析、手元のデータにより解析)では、解析対象例数がかかり減り、しかも偏った結果を得る可能性がある。そのため、多重代入法などにより欠測値を補完し解析することが推奨される。ただしその際、進行度などの観測値の欠測がどの欠測パターンで発生しているかを検討し、適切な代入法を選択することが重要である(Nur(2008))。Clark et al.(2003)らは生存解析において、予後因子は欠測値を含むことが多いため、その欠測値を補完することで完全事例解析よりも信頼性の高い結果が得られるとする一方、代入法による解析は欠測パターンが明らかな場合のみ適用できると限定している。欠測パターンに応じた欠測値補完により、より偏りの少ない結果が得られる。現状の資料を活用するためにも、適切な解析方法の開発が望まれる。

表4 仮想データ:多重代入法による補完例

ID	胸囲(cm)	肺活量(cc)	肺活量(cc) (欠測あり)	肺活量(cc) (補完値)
1	68	2000	2000	2000
2	71	1850	1850	1850
3	72	2100	2100	2100
4	72	1700	1700	1700
5	72	2200	2200	2200
6	76	2400	2400	2400
7	77	2150	2150	2150
8	77	2600	2600	2600
9	84	2300	.	3407
10	90	2800	.	2955

肺活量(補完値)は、多重代入法により補完した値である。

表5 多重代入法における結果の統合

	変数	係数	標準誤差	p 値
統合モデル	切片	-3366.19	2072.40	0.16
	胸囲	75.26	28.43	0.04
モデル 1	切片	-3408.17	828.83	<0.01
	胸囲	75.90	10.88	<0.01
モデル 2	切片	-1525.98	770.75	0.08
	胸囲	49.83	10.12	<0.01
モデル 3	切片	-1864.11	823.48	0.05
	胸囲	54.39	10.81	<0.01
モデル 4	切片	-4416.40	883.80	<0.01
	胸囲	89.80	11.60	<0.01
モデル 5	切片	-5616.31	1009.17	<0.01
	胸囲	106.37	13.25	<0.01

E. 結論

本報告では、悉皆性評価と欠測値の補完方法について概観した。地域がん登録の精度向上がすぐには期待できない現状においては、悉皆性評価、欠測値の補完を行うことで、より信頼性の高い資料活用が期待される。

F. 研究発表

なし

G. 知的所有権の取得状況

該当なし

H. 文献

味木和喜子, 津熊秀明, 大島明 (1998).
地域がん登録における登録の完全性の

評価指標およびそれを用いた大阪府がん登録の登録率の評価, *日本公衆衛生学雑誌*, 45(10), 1011-1016.

Brenner, H. (1995). Limitations of the Death Certificate Only Index as a Measurement of Incompleteness of Cancer Registration, *British Journal of Cancer*, 72(2), 506-510.

Bullard, J., Cloeman, M. P., Robinson, D., Luts, J. M., Bell, J. and Peto, J. (2000). Completeness of Cancer Registration: a New Method for Routine Use, *British Journal of Cancer*, 82(5), 1111-1116.

Buuren, S. V., Boshuizen, H. C. and Knook, D. L. (1999). Multiple

- Imputation of Missing Blood Pressure Covariates in Survival Analysis, *Statistics in Medicine*, 18(6), 681-694.
- Clark, T. G. and Altman, D. G. (2003). Developing a Prognostic Model in The Presence of Missing Data: an Ovarian Cancer Case Study, *Journal of Clinical Epidemiology*, 56(1), 28-37.
- Curado, M. P., Edwards, B., Shin, H. R., Storm, H., Ferlay, J., Heanue, M. and Boyle, P., eds. (2007). *Cancer Incidence in Five Continents, Vol. IX, IARC Scientific Publications No.160*, International Agency for Research on Cancer, Lyon. (Available at <http://www-dep.iarc.fr/>, 2009.2.7 accessed)
- Kamo, K., Kaneko, S., Satoh, K., Yanagihara, H., Mizuno, S. and Sobue, T. (2007). A Mathematical Estimation of True Cancer Incidence Using Data from Population-based Cancer Registries, *Japanese Journal of Clinical Oncology*, 37(2), 150-155.
- 加茂憲一 (2008). がん罹患の動向に関する研究, 厚生労働科学研究費補助金第三次対がん総合戦略研究事業「がん罹患・死亡動向の実態把握の研究」(主任研究者 祖父江友孝), 125-127, 平成19年度厚生労働科学研究費補助金総合報告書, 東京.
- Matsuda, T., Marugame, T., Kamo, K., Katanoda, K., Ajiki, W., Sobue, T. and Japan Cancer Surveillance Research Group. (2008). Cancer Incidence and Incidence Rates in Japan in 2002: Based on Data from 11 Population-based Cancer Registries, *Japanese Journal of Clinical Oncology*, 38(9), 641-648.
- Nur, U. (2008). Modeling Relative Survival in the Presence of Incomplete Data, *Handout of Cancer Survival Mini Sessions at Osaka Medical Center for Cancer and Cardiovascular*, Osaka.
- 祖父江友孝, 津熊秀明, 岡本直幸, 味木和喜子 (2007). *地域がん登録の手引き改訂第5版*, 地域がん登録全国協議会, 東京.
- Parkin, D. M. (2006). Standards to Ensure Quality of Cancer Registry Data, *JACR Monograph* (eds. Sobue, T. and Oshima, A.), 11, 1-17, Japanese Association of Cancer Registries, Osaka.
- Parkin, D. M., Chen, V. W., Ferlay, J., Galceran, J., Storm, H. and Whelan, S. L. (1994). Comparability and Quality Control in Cancer Registration, *IARC Technical Report*, 19, 18-19, International Agency for Research on Cancer, Lyon.
- Schmidtman, I. (2008). Estimating Completeness in Cancer Registries — Comparing Capture Recapture Method in a Simulation Study, *Biometrical Journal*, 50(6), 1077-1092.
- Schouten, L., Straatman, H., Kiemeny, L., Gimbrere, C. and Verbeek, A. (1994). The Capture-recapture Method for Estimation of Cancer Registry Completeness: A Useful Tool?, *International Journal of Epidemiology*, 23(6), 1111-1116.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.

Appendix.

1) Parkin の登録率推計式

仮定(1)より、登録率 r は以下のように展開される。

$$\begin{aligned} r &= \frac{a+b+c}{a+b+c+d} \\ &= \frac{a+b+c}{a+b+c+bc/a} \\ &= \frac{1}{1+\frac{bc/a}{a+b+c}} \\ &= \frac{1}{(1-x)+x(1+b/a)} \\ &= \frac{1}{(1-x)+\frac{x}{a/(a+b)}} \end{aligned}$$

更に仮定(2)、(3)より、

$$\begin{aligned} r &= \frac{1}{(1-x)+\frac{x}{(a+c)/(a+b+c)}} \\ &= \frac{1}{(1-x)+\frac{x}{1/y}} \end{aligned}$$

2) 味木の登録率推計式

一方味木らは、仮定(3)が日本における実情に一致しないことから、仮定(3)を用いない登録率の推計式を提案した。具体的には、

$$\begin{aligned} \frac{a}{a+b} &= \frac{(a+c)-c}{(a+b+c)-c} \\ &= \frac{(a+c)-x(a+b+c)}{(a+b+c)-x(a+b+c)} \\ &= \frac{1/y-x}{1-x} \end{aligned}$$

とし、これを Parkin の式に代入することで登録率の推計式を求めた。即ち、

$$\begin{aligned} r &= \frac{1}{(1-x)+\frac{x}{a/(a+b)}} \\ &= \frac{1}{(1-x)-\frac{x(1-x)}{1/y-x}} \\ &= \frac{1-xy}{1-x} \end{aligned}$$

上式が味木の式である。

3) 加茂の登録率推計式

加茂らは、がん罹患が全て DCN として把握された場合、IM 比は 1 になることに注目し、DCN と IM 比がともに 1 となる点を通る回帰曲線が存在すると仮定した。

次に、がん患者が発生するごとにすぐ登録が行われる地域がん登録を仮定すると、罹患数(真の罹患数) I 、IM 比(真の IM 比) β 、 a 、 b 、 c 、 d は、次の関係式をもつ。

$$\begin{aligned} a &= I/\beta, \\ b &= I(\beta-1)/\beta, \\ c &= d = 0 \end{aligned}$$

ここで、新たにがんと診断された患者 A が、登録漏れとなった場合、上式の関係は下記の通りになる。

$$\begin{aligned} a &= (I-A)/\beta, \\ b &= (I-A)/(\beta-1)/\beta, \\ c &= A/\beta, \\ d &= A(\beta-1)/\beta \end{aligned}$$

観察される罹患数は $a+b+c$ 、死亡数は $a+c$ であることから、IM 比 y と DCN% x は次式で示される。

$$\begin{aligned} y &= \frac{\beta I - A(\beta-1)}{I}, \\ x &= \frac{A}{\beta I - A(\beta-1)} \end{aligned}$$

上の 2 式を展開することにより、

$$y = \frac{\beta}{1+(\beta-1)x}$$

が得られる。

ここで、地域 i で観察される DCN% x_i 、IM 比 y_i とおく。 $y_i \in (1, \infty)$ から、次のモデルと式(4)は等しいと考えられる。

$$\tilde{y}_i = \log\left(\frac{\beta}{1+(\beta-1)x} - 1\right) + \varepsilon_i$$

ただし $\tilde{y}_i = \log(y_i - 1)$ であり、ランダムエラー ε_i が平均 0 分散 σ^2 の標準正規分布から独立であるとする。このモデルでの未知パラメータは β および σ^2 である。これらのパラメータは、人口で重み付けされた最尤推定量を最大にする値として推定する。特に σ^2 については Seber and Wild (1989) の報告が詳しい。これより、DCN% が 0 となるときの IM 比 (真の IM 比) を求め、それに観測されている死亡数を乗ずることにより、真の罹患数を推定する。

小児がんにおける治療の集中化の検討

研究協力者 筒井杏奈(大阪大学大学院医学系研究科)

研究代表者 大野ゆう子(大阪大学大学院医学系研究科)

地域がん登録は、その地域でのがん患者の罹患率と生存率だけでなく、患者の受療状況も収集している。今日、小児がん治療は疾患の希少性から集中化が必要だと言われているが、集中化指標がない為に実態把握が難しいという問題があった。

本研究では、地域がん登録で使用可能な集中化指標として、パレート分析を用いながら、治療医療機関数、パレート曲線とジニ係数、施設あたりの年平均治療件数の3手法を提案し、検討を行った。大阪府がん登録の集中化の動向をみたところ、少子化により患者数が減少した影響で治療医療機関数が減少した、集中化傾向に変化は見られなかった、欧米と比べて年平均治療件数が低かった、などの知見を得た。

集中化指標を用い、地域がん登録から治療の集中化の実態把握をすることは、医療政策に提言可能な基礎資料として、有効と考える。とりわけ、今日に急減している小児がんは正確な実態把握が急務である。今後、集中化指標について更なる検討を行い、また本指標を他の地域がん登録も適用して方法の妥当性を評価したい。

A. 研究目的

小児がんは15歳未満の子どもに発症した悪性新生物を指す。大人と比べて罹患率が大変低く、患者数が少ない。希少なために生存率の向上には、特定の医療機関に患者を集めて治療の技術集積を行うことが重要であると報告されてきた(Stiller(1988, 1994))。実際に英国等の欧米では、患者を特定の専門病院に集めて治療を行う集中化が行われている。たとえば英国では、一年間の罹患患者の90%にあたる1,350人を、22施設に集めて治療している(Ablett・Pinkerton(2003))。

治療の集中化について議論する際に、問題になるのが集中化指標である。地域がん登録の解析報告では「○%の患者が○施設に集中化している」という一文でしか集中化が表記されない。集中指標は数多く提案されたが、未だ集中の概念の定義すらないのが現状である(芳鐘(2000))。このため、地域

がん登録に患者の受療状況のデータがあっても、指標がない為に集中化の議論ができていない問題があった。

大阪府下の病院に勤務する専門医らによると、日本での小児がん治療については、1990年代までは個々の医療機関で行われ分散状態にあったものが1990年代に集中化したという。本研究では地域がん登録で使用可能な集中化の指標を幾つか提案し、実際の解析、評価を試みた。

B. 研究方法

本研究では提案した指標で集中化の有無や程度が把握できるか明らかにするために、地域がん登録を用いて解析を行った。本研究では地域がん登録のうち、小児人口が多く、推定登録率も高く(Ajiki・Hanai他(1998))、データ収集期間が長いという、主に3点の長所を持った大阪府がん登録で解析を行った。

大阪府がん登録に1975年から2002年までに登録された15歳未満の新発届出患者(n=5,291)のうち、上皮性腫瘍(n=3)、治療情報のない者(n=153)、他府県に所在する施設で治療を受けた者(n=417)を除外した。他府県に所在する施設で治療を受けた者を除外した理由は、どの施設で治療を受けたのかが明らかでないためである。最終解析対象者は4,738人であった。

対象者の罹患したがんの種類は、小児がんの国際分類に従って12の診断群に区分した(Kramarova・Stiller (1996))。また対象者を診断年により、1975年から1984年までの前期(n=1,976)、1985年から1993年までの中期(n=1,661)、1994年から2002年までの後期(n=1,101)の3期間に区分した。

全体の受療状況を把握するため、治療医療機関情報が明らかな者には治療医療機関を用い、不明の者には診断医療機関で代用した。診断医療機関と治療医療機関の両方が明らかな患者のうち、両者が一致していた患者は91.8%(3,435人)であった。

本研究では小児がん治療における集中化を「小児がん治療の専門家がおり、治療環境が整っている少数の医療機関に患者が集まるようになること」と考え、集中化に言及した先行研究で用いられていた指標をはじめいくつかの指標について検討した。

①治療医療機関数と患者数の経年変化

パレート分析より、上位50%群、上位75%群、全100%群の患者を治療している医療機関数を求め、変動を調べた。

②パレート曲線、ジニ係数の期間比較

期間別にパレート曲線、ジニ係数を求め、患者集中化の指標として検討した。ジニ係数は0から1までの値をとり、集中化するほど値が1に近くなる。なお、ジニ係数の値が同じでもパレート曲線の立ち上がり急峻な方が、集中化が進んでいると判断される。

③施設あたり年平均治療件数比較

パレート図において、期間別に、上位50%群、上位75%群、全100%群別の、施設あたりの年平均治療件数を求めた。今、当該観察期間*t*年間における小児がん罹患数を*m*、パレート図における上位*i*%までの患者治療を担う治療医療機関数を*X_i*とすると、年平均治療件数*z_i*は次式のように示せる。

$$z_i = \frac{m \cdot i}{100 \cdot t \cdot X_i}$$

上位90%群についても年平均治療件数を求め、報告があったイギリス、ドイツ、スイスおよびアメリカ(ジョージア州)の値と国際比較を行った((Ablett・Pinkerton (2003), Howell・Ward 他 (2007), Kaatsch・Spix (2004), Swiss Childhood Cancer Registry (2005))。

(倫理面への配慮)

大阪府がん登録資料利用適否の審査と承認を得た上で、研究に必要なデータの提供を受けた。本研究は患者を特定する項目は含まない。

C. 研究結果

①治療医療機関数と患者数の経年変化

大阪府の小児人口は1975年には212万人であったが、2002年には123万人に半減した(図1)。この期間中は常に減少傾向にあり、特に1980年から1985年にかけて急減した。小児がん患者数は1975年から1990年まではほぼ変化が見られなかったが、1990年頃から急減し、1975年の185人から2002年には107人にまで半減した。

治療医療機関数の変化を表1に示した。上位50%群、上位75%群、全100%群の治療医療機関数は1975年から2002年にかけて減少していた。上位50%群の診断医療機関数は前期では7、中期は6、後期は3であり、上位75%群の前期は15、中期は15、後期は8、全100%群の前期は91、中期は86、後期は56であった。3群共に前期から中期にかけては微減もしくは変化せず、中期から後期にかけて2分の1程度の減少が見られた。

診断群別に見ると(表1)、Ⅷ.悪性骨腫瘍の上位50%群を除き、上位50%群、上位75%群、全100%群において後期の治療医療機関数が前期と比較して減少した。後期における治療医療機関数が最も少ないのは1施設のⅣ.交感神経系腫瘍、Ⅴ.網膜芽細胞腫であり、最も多いのはⅠ.白血病、Ⅱ.リンパ腫・細胞内皮系腫瘍、Ⅲ.中枢神経系腫瘍、Ⅹ.性腺・胚細胞腫瘍だった。

また、年平均罹患数を見ると、小児がん全体で前期は197.6人、中期は184.6人、後期

は122.3人と減少傾向にあった。診断群別に見ても、IV.交感神経系腫瘍を除き、年平均罹患数は前期から中期にかけては若干の増減が見られ、中期から後期にかけては大きく減少していた。後期では12群中8群の年平均罹患数が、10人以下であった。例外的に、IV.交感神経系腫瘍は前期に119人だった罹患数が中期には206人に倍増し、後期でも160人と高い値を示した。

②パレート曲線、ジニ係数の期間比較

期間別の治療機関数を総数としてパレート曲線とジニ係数を求めた。パレート曲線は、3期間とも変動がなかったが、後期にやや左上に移動していた。ジニ係数は3期間を通じて変動が少なかった(図2)。

診断群別にジニ係数の値を比較すると、前期の平均値は 0.504 ± 0.165 、最小値は0.382、最大値は0.642、中期の平均値は 0.499 ± 0.206 、最小値は0.275、最大値は0.657、後期の平均値は 0.453 ± 0.242 、最小値は0.275、最大値は0.660であった(図3)。より希少な診断群、すなわちV.網膜芽細胞腫、VI.腎腫瘍、VII.肝腫瘍、VIII.悪性骨腫瘍、IX.軟部組織肉腫、XI.上皮性腫瘍およびXII.その他及び不詳ほどジニ係数が低かった。

③施設あたりの年平均治療件数比較

年平均治療件数の変化をみると、総数では3期間通して2.2人であった(前期:2.17人、中期:2.15人、後期:2.18人)。上位90%群で見ると、前期は5.6人、中期は5.5人、後期は6.1人であり、上位50%群では、前期は14.1人、中期は15.4人、後期は20.4人であった。

諸外国の値をみると、上位90%群において、英国は61.4人、米国ジョージア州は60.9人、ドイツは32.4人、スイス27.1人であった。

D. 考察

本研究より、1993年から1994年にかけて、上位50%群、上位75%群、全100%群における治療医療機関数の3分の1程度の顕著な減少が観察された。大規模医療機関の新設および移転によって大阪府内の小児がんの医療体制が変化し、治療医療機関数にも影響を与えたと推察される。

大阪府では小児人口も検討対象期間で

減少しており、この少子化の影響で小児がんの罹患数が28年間で2分の1程度に減少したことが観察された。また、治療医療機関数の減少が、小児がんの罹患数の減少と同時に起きていたことが観察された。

治療医療機関数だけをみれば集中化が進んだように見えるが、患者数の減少により単純には結論がでないことが示された。大阪府下の小児診療科数が約2分の1と25年間で大きく減ったとは考え難いため、紹介だけを行うようになった医療機関もあることは推察される。しかし、がん登録においては治療せずに紹介した医療機関まで把握できていないため、数の変化からだけではこれ以上の検討は難しいと考えられる。

パレート曲線は、3期間中にほぼ同じという結果となった。診断群別ではより罹患率の高い診断群ほど、他の診断群と比較して治療が集中していることが、パレート曲線とジニ係数から確認された。原因として、1990年代から多施設共同型の臨床試験が小児がんでも実施されるようになったことが考えられる(牧本(2004))。臨床試験は特に罹患率のより高いI.白血病やII.リンパ腫・細胞内皮系腫瘍などの群で主に行われたため(牧本(2004))、より罹患率の高い診断群ほど集中化の傾向が見られたと思われる。IV.交感神経系腫瘍では前期から中期にかけて集中化せず、後期に集中化していたことが確認された。これは、1981年に一部地域で試験的に開始され、その後1985年から2003年にかけて府全体を対象に行われた、神経芽細胞腫瘍マスキングの影響と考えられる(Ajiki・Tsukuma他(1998))。なお、このスクリーニングは結果的に過剰診断が指摘され2003年に廃止されている(神経芽細胞腫瘍マスキング検査のあり方に関する検討会(2003))。

なお、ジニ係数は標本数(施設数)が50以下の場合、過小評価バイアス、すなわちより低く算出される性質が知られているが(George(2003), Lerman・Yitzhaki(1989))、診断群別パレート図に於ける検討でも後期における集中化が確認できなかった診断群も多いため、集中化傾向は明確にはみられなかったという結論は同じと考えられる。

希少がんが多い小児がんの場合、ジニ係数は補助的な指標としては有効ではあるがパレート図による検討が望ましいこと、ただし治