

## 病院の患者データに基づく有病率の推定法 - 通院距離 - 患者数関係の解析 -

スライド1のように、どのような疾患がどの程度の頻度で発生しているかということ把握することは、我が国の保健医療を考える上で非常に重要なことであると考えます。しかしながらこうした値を知るためには、ある調査区域・地区を設定して、その当該地区での全医療機関の協力を得て調査するという方法を探らざるを得ません。しかし、このような調査をすると非常に多くの時間が必要で、また労力もかかるということが問題です。

一方我々阪大病院では大阪の北摂地区を中心として、1日約2,000人の患者さんが来院します。我々は1993年に、医師が直接診断名を登録するシステムを導入して、本院では通院患者について、どのような患者さんがどれくらいの頻度で来ておられるかということ、かなりの精度で把握することができるようになっています。しかし、我々のような大学病院の性格上、恐らく治療の難しい疾患については遠方であっても来られるでしょうし、逆に比較的簡単に治療できる疾患については、わざわざ遠方から来られることはまれだろうと推測されます。従って、疾患の性質によって我々のような一病院に集まる患者さんの集積の仕方が影響を受けますので、いくら我々の病院でどの疾患の患者がどのくらいの数来られるかということ正確に把握したとしても、それをもって全国の期間有病率あるいは罹患率を推定するということはできません(スライド2)。

スライド1

ある疾患の有病率(または罹患率)を知る

- 調査地区の設定
- 当該地区での全医療機関における調査

しかし、

- 必要とされる時間、労力が問題。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド2

院内統計値と大規模調査による統計値

院内統計 ≠ 大規模調査による統計

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド3

目的

院内データから有病率、罹患率を推定する手法の確立

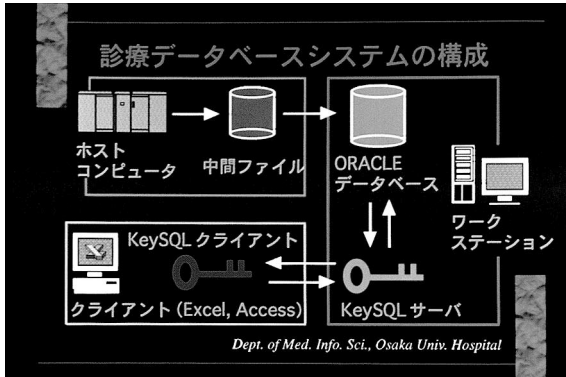
通院患者の住居分布の解析を行い、施設特有のバイアスを補正することにより、統計値を推定する。

Dept. of Med. Info. Sci., Osaka Univ. Hospital



松村 泰志 先生  
大阪大学医学部附属病院  
医療情報部助手

スライド 4



スライド 5

対象患者  
1995年4月1日～1996年3月31日に本院を外来受診した患者

対象疾患統計値  
①悪性新生物(がん)の主要11部位がん罹患率  
②特定疾患(いわゆる難病)のうち、比較的頻度の高い9疾患の有病率

について、診療データベースから得られた患者データに基づき、推定を行う

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 6

データ抽出  
①当該期間に新規登録された患者  
②当該期間にその病名がアクティブであった患者  
ただし、①②とも保険病名、疑い病名は除く

比較対象  
大阪府環境保健部による調査結果  
①「大阪府におけるがん登録(1992)」  
②「在宅難病患者訪問指導事業実施状況一覧(1994)」

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 7

患者分布モデル  
有病率  $\alpha$  が地域によらず一定とすると、

$$\alpha \cdot p_i = n_i / N_i \quad \dots [1]$$

が成り立つ。ただし、  
 $p_i$  : 地域  $i$  の疾患 A 患者が本院を受診する確率  
 $n_i$  : 地域  $i$  から本院を受診した疾患 A 患者数  
 $N_i$  : 地域  $i$  全人口

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 4 で本院のデータベースの構造を少しご紹介します。

本院では患者さんのデータは全てホストコンピューターに蓄積されていますが、これをそのまま解析に使うと結構手間がかかります。そこで、この中で解析の価値のあるデータを抽出して、ワークステーションの方に毎日夜間バッチ処理でダウンロードしています。現在のところ患者基本情報、外来・入院の診療履歴、診断名、薬等の情報がデータベースに蓄積されています。これをKey SQLというミドルウェアを介してクライアントマシン(パーソナルコンピューター)で必要なデータを検索・抽出して、さらにExcelとかAccessという市販のソフトを使って解析を行ないます。PCの使い方に少し詳しい方であれば、誰でも病院のデータが解析できるという環境を整えています。

スライド 5 は今回の研究の対象とした患者さんを示しています。1995年4月1日から1996年3月31日に本院を外来受診した患者さんを抽出しています。対象疾患は、データが既知であるということが必要でしたので、悪性新生物の主要11部位の癌、それから特定疾患 いわゆる難病のうちの比較的頻度の高い9疾患について調べました。

(スライド 6) 癌は当該期間に新規登録された患者さんを抽出しました。そのデータを、大阪府環境保健部調査による1992年のデータが比較的詳しく出ていましたので、これを真値として、我々の解析で求めた値と比較しました。特定疾患については、その当該期間にその病名がアクティブであったものを抽出の対象としました。そのデータを同じく大阪府の調査による在宅難病患者の一覧表から得られたデータと比較しました。

まず、考え方を示します(スライド 7)

有病率( )を地域によらず一定とすると、ある地域  $i$  における、その地域から本院を受診された患者数 ( $n_i$ ) の地域の人口 ( $N_i$ ) との比(これを患者人口密度という言葉で表現します)は、有病率 にその地域において

その疾患の患者さんが阪大病院を受診する確率 ( $p_i$ ) を掛けた値に等しいということが成り立ちます。

スライド8に示すとおり、この患者人口密度とその地区の市役所、あるいは区役所と本院との距離を  $d_i$  として、これをプロットしました。

スライド9は胃癌の患者さんについてプロットしたものです。横軸は通院距離で、縦軸は患者密度を片対数で表わしています。この図で見ていただいでわかるように、だいたい15 km位までの間は比較的きれいな直線上に乗ります。すなわち指数関数であてはめることができるということです。

(スライド10)つまり、患者密度はCに指数減衰関数を掛けたものという式で表わすことができ、この左辺(人口密度)は先ほどの式から  $\alpha \cdot p_i$  とおくことができます。ここで距離  $d$  を本院にどんどん近付けていくと、基本的に1になるだろう すなわち阪大のごく近辺においては皆さん阪大病院に来るだろうという仮定をおくと、この式から  $\alpha = C$  になります。すなわち先ほどのグラフのy切片がちょうど有病率あるいは罹患率ということになります。

ここで、この理論を使って実際にデータを求めていきます。データ数が多ければ多いほど良いわけですが、遠方の地域のデータは直線から離れてしまうという性質がありますので、慎重にアプローチしました。スライド11のように本院を含む近隣5都市についてまずデータを求めました。この定数項(y切片)を求めます。それから順次阪大病院に近いものを1つずつ加えていき、同様の計算をしていきます。その計算を距離が15 kmになる都市まで続けていくわけですが、その中で回帰曲線の  $R^2$  値が最大であって、かつF値が1%以下で有意

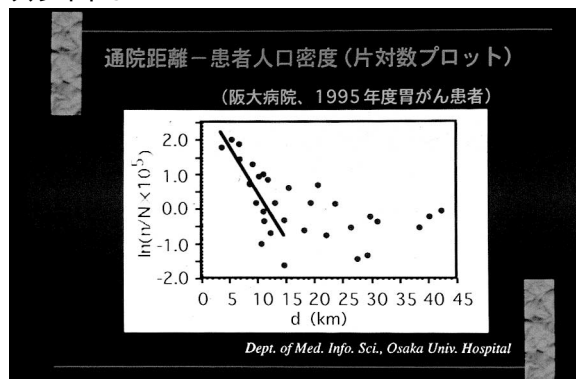
スライド8

市(または区)行政区分ごとに、患者数  $n_i$  を求め、人口数  $N_i$  との比 ( $n_i/N_i$ ) を算出する。  
 これらの値を、本院と市(区)役所の距離  $d_i$  に対してプロットする。

$n_i/N_i$ : 地域  $i$  の患者人口密度  
 $d_i$ : 地域  $i$  患者の通院距離

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド9



スライド10

$n/N = C \cdot \exp(-\beta \cdot d)$  ( $d = 0 \sim 15\text{km}$ ) ...[2]

よって、[1]式  $\alpha \cdot p_i = n_i/N_i$  から

$\alpha \cdot p(d) = C \cdot \exp(-\beta \cdot d)$  ...[3]  
 ( $p_i \rightarrow p(d)$ )

本院のごく近隣では、疾患Aに罹患した患者がすべて本院を受診すると仮定すると、

$p(d) \rightarrow 1$  ( $d \rightarrow 0$ )  $\Rightarrow$   $\alpha = C$  ( $d \rightarrow 0$ )

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド11

手順 1

通院距離 - 患者人口密度の片対数プロットにおいて、

- 本院所在都市を含む近隣5都市(吹田市、豊中市、箕面市、茨木市、摂津市)について、人口重みつき回帰直線の定数項値(y切片:  $y_0$ )を求める。
- 都市を順次1つずつ、本院からの距離が近い順に加えて、同様に定数項値  $y_0$  を求め、距離が15kmになるまで続ける。
- 以上より得られた値のうち、決定係数  $R^2$  が最大かつF値が1%有意となるものを推定値  $E$  とした。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 12

### 手順 2

- 隣接 5 都市を A ゾーン、A ゾーン以外で距離 15km 以内の都市を B ゾーンとし、ゾーンごとに患者人口密度を求める。
- 手順 1 と同様に片対数プロットを行い、得られた 2 点を通る直線式の定数項値 (y 切片:  $\hat{y}_0$ ) を求める。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 13

- 手順 1 で推定値が求められた疾患について、この定数項値  $\hat{y}_0$  と手順 1 による推定値のプロットを行い、最小自乗法により予測式を求める。
- 手順 1 で推定不可能であった疾患に予測式を適用し、手順 2 による推定値  $\hat{E}$  とする。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 14

### 推定値と大阪府統計値の比較

#### がん罹患率(1)

	本院患者数	大阪府 (T)	推定値 (E)	E/T	R <sup>2</sup>	手順
食道がん	88	7.7	5.7	0.74	-	2
胃がん	171	57.7	35.6	0.62	0.673	1
結腸がん	70	25.3	9.5	0.38	-	2
直腸がん	45	13.2	3.0	0.22	-	2
結腸+直腸	115	38.4	25.8	0.67	0.563	1
肝臓がん	158	37.6	21.9	0.58	-	2

単位：本院患者数は [人]、E および T は [10 万人対罹患率]

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 15

### がん罹患率(2)

	本院患者数	大阪府 (T)	推定値 (E)	E/T	R <sup>2</sup>	手順
胆嚢がん	30	9.0	1.6	0.18	-	2
膵臓がん	46	10.9	4.3	0.39	-	2
肺がん	136	39.1	44.2	1.13	0.571	1
乳がん	189	16.1	48.2	3.00	0.675	1
子宮がん	135	11.3	26.9	2.38	-	2
前立腺がん	71	4.2	11.0	2.63	0.654	1

単位：本院患者数は [人]、E および T は [10 万人対罹患率]

Dept. of Med. Info. Sci., Osaka Univ. Hospital

であるものが得られた場合に、それを推定値としました。

しかし、本院受診患者数がそもそも非常に少ない場合では、こうした解析を行なうとかなりバラ付いてしまいます。そこでそういうケースでも何とか値を求められるように、スライド 12 の手順 2 を考えました。

この手順 2 は隣接 5 都市を全部ひっくるめて A ゾーン、それから 15 km 圏内かつ A ゾーン以外の都市を全部まとめたものを B ゾーンとして、それぞれの患者人口密度を求めました。それからその人口比の重みをつけて A ゾーン、B ゾーンを中心地区を求めて図のようにプロットしました。そして、この線の y 切片をまず求めます。スライド 13 に移ります。

この手順によって求めた y 切片と手順 1 で求めた E の間で最小自乗法で直線を引いて、この関係式を求め、この関係式で先ほどの手順 2 で求めた y 切片から、推定値 E を求めるという方法を用いました。

スライド 14 が結果です。まず癌患者のリストですが、例えば胃癌ですと本院で 171 例ありましたが、大阪府では 57.7 (10 万人対です) 人で、本院の先ほどの方法で推定した値が 35.6 人ということで、この推定値の真値との比率は 0.62 となります。というようなことをずっとやっていきました。

スライド 15 がその他の結果です。乳癌のようにかなり離れてしまうものもありますが、比較的近い値となっています。ただ、やはり患者数が少ないものについてはかなり小さめに出るという傾向があります。

スライド 16

特定疾患有病率(1)

	本院 患者数	大阪府 (T)	推定値 (E)	E/T	R <sup>2</sup>	手順
全身性エリテマトーデス	205	18.8	27.9	1.48	0.628	1
パーキンソン病	144	11.6	39.7	3.42	0.565	1
潰瘍性大腸炎	121	21.1	32.5	1.54	0.692	1
特発性血小板減少性紫斑病	113	12.2	15.8	1.30	0.662	1

単位：本院患者数は[人]、EおよびTは[10万人対相対有病率]

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 17

特定疾患有病率(2)

	本院 患者数	大阪府 (T)	推定値 (E)	E/T	R <sup>2</sup>	手順
ベーチェット病	96	7.0	17.0	2.43	-	2
クローン病	89	6.0	6.6	1.10	-	2
強皮症、皮膚筋炎及び多発性筋炎	48	8.4	7.6	0.91	-	2
ピュルガー病	31	6.0	3.7	0.62	-	2
脊髄小脳変性症	27	6.0	2.2	0.36	-	2

単位：本院患者数は[人]、EおよびTは[10万人対相対有病率]

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 18

大阪府統計値 (T) に対する推定値 (E) の比 (E/T)

- 中央値 0.91
- 95%信頼区間 0.58 - 1.54

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 19

まとめ

- 通院距離-患者人口密度のプロットは、病院の近隣地域において、指数減衰関数によって近似可能であった。
- われわれのモデルによれば、有病率および罹患率の推定値は、近似曲線の定数項値  $y_0$  に等しい。
- 本法により得られた有病率、罹患率の推定値は、大阪府による調査結果とオーダレベルでの一致がみられた。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

スライド 16 は特定疾患についての結果ですが、先ほどと同様な傾向があります。

スライド 17 は残りの特定疾患です。

以上の結果を全てまとめて評価しますと(スライド 18)、中央値 0.91 と真値より若干低めに出るものの、non-parametric の方法で 95% 信頼区間を求めたところ、0.58 ~ 1.54 と、この方法でもそこその精度で推定ができることがわかりました。

スライド 19 はまとめです。我々は通院距離 患者人口密度のプロットを用いて、これが指数減衰関数によって近似できるということを示しました。これによって定数項  $y_0$  を求め、それで有病率あるいは罹患率が推定できるかについて検討しました。その結果オーダレベルでの一致が見られました。

以上より(スライド 20)、本院のような比較的センター病院の機能を持つ病院のデータから、ある程度の精度をもって有病率あるいは罹患率の推定が可能であるということが示されました。

質疑応答

Q：非常に興味深く伺ったんですけども、距離ゼロであった場合に阪大病院に行く人の数が患者数の実態を反映しているという大前提があるように思えるんですけども。

スライド 20

結論

本法により、一センター病院のデータから、ある程度の精度をもって、有病率および罹患率の推定が可能であることが示された。

Dept. of Med. Info. Sci., Osaka Univ. Hospital

実際のところ、例えば大阪のように交通機関が均質に発達していて、同じ距離に住んでいることが同じ意味を持つといったところだったらいい。しかし、地方のように一本しか交通機関が無いところだと、それに沿った方向では例えば10%の人が来るとしても、逆方向だと全然違ってくるといようなことがあります。こうした地域においては、これはまず成立しないのではないかという疑問。

それと、距離ゼロであったときに阪大病院に受診される率に対する影響因子をもう少し...例えば教育程度であるとか年齢層であるとか、そういった影響因子を明らかにすることによって、より正確な予想が可能になるのではないかということについて、お答えをお願いします。

A: 先生がおっしゃる通りなんです。我々は、スタディを簡単にするために、地図上の直線距離をもって横軸を表わしてしまったのですが、おっしゃる通りよく見ると、例えば京阪沿線地区のほうから来られる患者は少ないですし、少し遠くても阪急電車を使って来られる患者は多いとかですね、交通手段による影響というのは非常に大きく受けております。

従いまして、どちらかと言うとやはり通院時間で横軸をとった方が良かったかなと反省しているわけですが、しかしながら、そう心配しながらも、実際のデータが割合に見事に直線に乗るところに、私も実は驚いたところです。それぞれの患者さんは、自分の判断でどの病院を受診するかということを考えておられるのですが、マスで見た場合にはきれいに距離と関係してしまうという特徴がございます。

それを逆に利用したというのが、今回の検討なんです。先生おっしゃるようにもう少し緻密に評価していけば、更に高い精度で評価できるのかもしれない。そのへんはこれから検討させていただきたいと思っています。

Q: 阪大病院というのもまた、ひとつの性格がわかるような気がするんですね。そういう意味で確かに先生は1つの病院をとられたけども、いくつかの病院をとって、観測地点をたくさん出せばもう少しはっきりするでしょうけども。

ただ、いずれにしても、これが何の役に立つのだろうか。そのこと自身は面白いんですけども、やっぱり僕は、病気は医者がつくるものだということを忘れちゃならないと思うんです。

私も臨床やってきました。私が山形に行ったときに頸部脊髄症という病気が無いんですよ。無いのは実は診断できる医者がいなかったからということなんです。その後だんだん増えてきた。こういうことはいくらでもあるわけですね。

ですからある病気の保健病名からとって、それをいくら考えてみても、どれくらい多いかということとは、やはりそういうことを診断する医者がどのくらいいるのか 何を専攻する医者が何人いるかということなんです。

そうすると、地域を決めて、各科の認定医なんかを複数揃えて、ぱっと診察すると、まるで違う数字が出てきますよ。これはさっき関節の研究でご発表された先生が言われていた「高血圧と比べて関節症が50%で、もう一つが30%」 そんな数字を誰が信じています。僕は信じるんですよ。知らないんですよ、皆さん。

ですから何を病気とするのかということで、ではこれを何に利用するかということを考える場合に、ある一定の何かのときには使えるんだけど、実際にこのデータを用いて、じゃあどうい医療にどういふうに資源を配分するかなんてことに利用されたらたまったもんじゃないし、やっぱり手間がかかって地域を決めて、非常にやり方が難しいのですけれど、試験してみる。

実は私ちょっとやったことがあるんです。ただ、整形外科医だけでやったもので、色々バイアスが入ってくるでしょ。まるで違うんです。で、われわれも臨床やって見るのと、地域に出て行ってそういうのをやってみると、こんなに違うものかと実はびっくりしたんです。ですからまあ、その使い方と言うか、それを十分気を付けて...やはり病気というのは医者を作るんだ。そういう面がものすごいあるんだと

いうことを、やはり忘れてはいけない。先生最初言われましたよね、でもそれを忘れちゃいかんと、ちょっと感じましたので付け加えさせていただきました。

A: 先生がおっしゃる通りだと思います。今回比較的成功いたしましたのは、選んだ疾患が比較的良かったからです。つまり癌にしましても特定疾患にしましても、やはりきちっと病名登録しておかないと、色々な不都合が生じますから、医者の方も真面目に入れるだろうと予想されます。

しかしながらこれを他の疾患全てについて適応できるかという、私もそれについては少し悲観的に考えております。ただ、そうは言いますが、ここでこうした数式で比較的良い値が出るということになりますと、一大学病院で、例えば心移植の対象となる患者さんがどれくらいいるかということ进行调查いたしまして、かつその住所も調べておく。そうしますと本来は、本院に何人いるから恐らく全国に何人いるだろう、というようなことを言うのは非常に難しいわけですが、しかしそこに住所の情報も加えて今の式に当てはめて評価すると、一つの参考値が出てくるわけですね。

というように、調査のほうはもう少しきちっと...決して病名登録のデータを信じるのではなく、きちっとした評価でもって定義された疾患をちゃんとカウントする。その上でこの理論的方法を使って、何かそういう全国推定に使えるのではないかと、こういうふうを考えてこのスタディを行っております。

Q: ほんのわずかですが精神疾患の疫学をしている立場から、質問というよりもコメントさせていただきたいと思えます。

技術的なことについてはいくつかいちゃもんをつけたい部分も多々あるんですが、それは置いておきまして、大変感心いたしましたのは、こういう点です。

疫学調査は先ほどのご質問の方がおっしゃられたように、直接面接で診断を確定して地域を決定してというのは大原則ですし私共もやっております。しかしながら例えば私共の領域で考えますと、非常に小ぶりの疫学調査をしても数百万円実費がかかりますし、中ぶりの疫学調査を行なうと数千万円、アメリカやヨーロッパのデータと負けられないようなものをやろうとすれば(そんな大がかりのことはやったことはございませんが) 実費計算しますとだいたい1億円かかります。そうしますと、ある特定の疾患について疫学調査をするといっても、科研費であろうと、厚生科学研究費、科技厅の研究費、他の各取材だの研究費であろうと、「すみません、この疾患の疫学調査をするので8,000万いただきます」と言っても出すところは無いんです。

例えば私は地域限定型の、ものすごく小さな地域で、お恥ずかしいんですが、住民500人ある町を限定して精神疾患の障害有病率を求めることをやっておりました。そこで、癌の研究者はいいなと思いたしたのは、癌の患者さんというのはまず90何%最終的には受診なさる。それを前提に考えますと、今の先生のような仮説が立てられて、そしてきれいなデータでその仮説通りの推定有病率が出るということ。これはそういう意味では、何学と言って良いのかわかりませんが、疫学調査経済学みたいな観点からしますと、大変素晴らしい研究だと思いましたので感銘いたしました。ありがとうございました。