



Title	Estimation of disease incidence and prevalence using a single dataset from the data warehouse of a hospital information system
Author(s)	桑田, 成規
Citation	大阪大学, 2003, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/43982">https://hdl.handle.net/11094/43982</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏名	桑田成規
博士の専攻分野の名称	博士(医学)
学位記番号	第17697号
学位授与年月日	平成15年3月25日
学位授与の要件	学位規則第4条第1項該当 医学系研究科内科系専攻
学位論文名	Estimation of disease incidence and prevalence using a single dataset from the data warehouse of a hospital information system (病院情報システムのデータウェアハウスから抽出した単一データセットに基づく罹患率および有病率の推定方法)
論文審査委員	(主査) 教授 武田 裕 (副査) 教授 多田羅浩三 教授 川瀬 一郎

### 論文内容の要旨

#### 〔目的〕

罹患率や有病率などの疫学指標を算出するためには、一般に、調査対象地域の悉皆調査が行われる。その調査には多くの労力と費用が必要とされるため、より簡便な方法で疫学指標を推定する手法に対するニーズが存在する。

近年、病院情報システムを導入する医療施設が増え、来院患者の疾患頻度を容易に知ることができるようになった。しかし、一医療施設でのデータは、施設周辺地域の疾病構造を正確に反映しているわけではない。すなわち、医療施設における患者の集中度が疾患ごとに異なるため、来院患者の疾患頻度から、周辺地域の罹患率および有病率を推定することは難しい。しかし、来院患者の住居分布を調べることにより、疾患ごとの患者集中度によるバイアスを除き、周辺地域における罹患率および有病率が推定できる可能性がある。

本研究では、院内統計値から周辺地域の疫学指標を推定する方法を確立することを目的とする。悪性新生物(以下、癌)の罹患率、および特定疾患(以下、難治疾患)の有病率が大阪府環境保健部ほかの調査により求められている。これらの疾患を対象として、本法の妥当性を評価した。

#### 〔方法ならびに成績〕

##### 1. データ抽出方法

大阪大学医学部附属病院(以下、本院)病院情報システムのデータウェアハウスから対象となるデータの抽出を行った。

##### 2. 対象

###### ①癌罹患率の推定に用いられた患者データ

1998年4月1日から1999年3月31日までに新規病名登録された、主要11部位(食道、胃、直腸、結腸、肺、肝臓、脾臓、胆嚢、乳房、子宮、前立腺)の癌患者を解析の対象とした。

###### ②難治疾患有病率の推定に用いられた患者データ

1998年4月1日から1999年3月31日までに本院を外来受診した患者のうち、比較的頻度の高い9つの難治疾患

(全身性エリテマトーデス (SLE)、パーキンソン病、潰瘍性大腸炎、特発性血小板減少性紫斑病、ベーチェット病、クローン病、強皮症・皮膚筋炎および多発性筋炎、ビュルガー病、脊髄小脳変性症) に罹患している患者を解析の対象とした。

### 3. 患者分布モデル

ある疾患 X について、地域  $i$  から本院を受診した疾患 X 患者数を  $n_i$ 、地域  $i$  人口を  $N_i$ 、地域  $i$  の疾患 X 患者が本院を受診する確率を  $p_i$  とする。罹患 (有病) 率  $\alpha$  が地域によらず一定と仮定する場合、次の式が成立する。

$$\alpha p_i = n_i/N_i \quad (1)$$

ここで(1)式右辺 ( $n_i/N_i$ ) を患者人口密度と呼ぶ。次に、市・区行政区画を単位として本院周辺地域を分割し、実際にいくつかの疾患について、患者人口密度 ( $n_i/N_i$ ) と、地域  $i$  から本院への平均通院距離 ( $d$ ) の関係を調べたところ、本院の近傍地域 ( $d < 10 \sim 15 \text{ km}$ ) において、患者人口密度は平均通院距離の指數減衰関数として近似可能であることがわかった。

$$n/N = B \exp(-\beta d) \quad (B \text{ および } \beta \text{ は定数}) \quad (2)$$

したがって、式(1)(2)より、

$$\alpha p = B \exp(-\beta d) \quad (3)$$

が成立する。ここで、本院のごく近隣地域において、疾患 X に罹患した患者はすべて本院を受診する、すなわち、 $d \rightarrow 0$  のとき  $p \rightarrow 1$  と仮定すると、式(3)は、

$$\alpha = B \quad (d \rightarrow 0) \quad (4)$$

となり、罹患 (有病) 率は、指數減衰関数の定数項  $B$  に等しい。

### 4. 計算方法

3. で示した患者分布モデルに従い、2. ①②の患者データを用いて、癌罹患率および難治疾患有病率の推定を行った。両データとも適用される手法は共通である。

(手順 1) 本院の所在都市を含む隣接 5 地域について、平均通院距離 - 患者人口密度の片対数プロットを行い、人口重みつき回帰直線を求めた。次に、5 地域以外で最も距離の近い 1 地域を加えて再度回帰直線を求め、順次、通院距離 15 km 以下の地域がすべて含まれるまで同様の手順を繰り返した。以上より得られた回帰直線の定数項  $B$  のうち、決定係数  $R^2$  が 0.6 以上で、極大値をとり、かつ 5 % 有意となるものを推定値として採用した。これを手順 1 とする。

手順 1 により推定値が求められない場合には、次の手順 2 を適用した。

(手順 2) 通院距離 15 km 以内の地域の全プロット ( $n$  個) に対し、ある 1 点を除いた ( $n-1$ ) 個のプロットで直線回帰を行った場合に、その  $R^2$  が最大値となるプロット群を求めた。この操作を再帰的に繰り返し、順次プロット数を減じた。繰り返し回数  $k$  に対し、 $R^2$  の増分を  $\Delta R^2(k) = R^2(k) - R^2(k-1)$  ( $k = 2, 3, \dots$ ) と定義するとき、 $\Delta R^2(k+1)/\Delta R^2(k)$  が 0.5 以下、かつ  $\Delta R^2(k)$  が基準値を超えた時点でプロットの除去を中止した。以上の操作より得られたプロットに対し、近接 5 都市を隣接 5 都市とみなして、再び手順 1 を通用した。これを手順 2 とする。

### 5. 結 果

直腸癌、胆嚢癌、膵臓癌、クローン病、およびビュルガー病については手順 2、これ以外の疾患については手順 1 により推定値が求められた。ただし、脊髄小脳変性症については、対象患者数が少ないうえプロットのばらつきが大きく、手順 1、2 のいずれにあっても有意な結果が得られなかつた。

本法による推定値の妥当性を検討するため、大阪府による統計調査結果（府内における癌罹患者数（1998 年）、特定疾患有病者数（1995 年））との比較を行った。本法による推定では、推定値が得られたすべての疾患で、大阪府統計値が推定値の 95 % 信頼区間に入り、良好な一致がみられた。本院患者数が少ない疾患、およびプロットのばらつきが大きい疾患については、推定の精度が低い傾向がみられた。

本院患者数と大阪府統計値との相関係数は  $R=0.589$  であったのに対し、本法による推定値と大阪府統計値との相関係数は  $R=0.934$  と良好であった。

## [ 総括 ]

一医療施設における来院患者データを用いて、当該施設周辺地域における罹患率および有病率を推定する方法を確立した。

本法では、患者の住居分布を解析し、患者集中度によるバイアスを除くことで、罹患率および有病率の推定を行った。得られた結果と大阪府による調査値を比較した結果、両者の間には良好な相関関係が認められた。

本法により、大規模調査と比べてより少ない費用と労力で、ある程度の精度をもった推定が可能であることが示された。

## 論文審査の結果の要旨

本研究は、一施設内で得られる患者データのみを用い、データマイニング手法を応用して、当該施設の周辺地域の罹患率および有病率を推定する方法を確立したものである。推定値を求めるに当たっては、疾患ごとに患者集中度が異なることによるバイアスをどのように除去するかが課題となるが、本研究では、患者の来院確率（地域人口あたりの来院患者数）が来院患者の通院距離に対して指數関数的に減衰するという患者分布モデルを構築することにより、この課題を解決した。さらに、患者の通院距離に補正を加える手法、および地域人口により重みを付けて回帰直線を求める手法を採用することにより、推定の精度をさらに高める工夫を行った。本研究では、大阪大学医学部附属病院内に構築されたデータウェアハウスを利用し、当院の患者データに基づいて、主要 11 部位の悪性新生物の罹患率、および比較的頻度の高い 9 つの特定疾患の有病率を推定した。その結果、得られた推定値を大阪府による調査統計値と比較したところ、全 20 疾患のうち 19 疾患について、大阪府統計値が推定値の 95% 信頼区間に位置すること、および推定値と大阪府統計値の間には高い相関が見られることから、両者に良好な一致が認められると判断された。

本研究により提示された方法は、簡便な調査による臨床疫学調査法として独創的であり、また新たな病院機能評価方法としても適当であることから、学位に値するものと認める。