
一般口演

一般口演6

DPC・地理情報・地域連携

2017年11月21日(火) 08:30 ~ 10:00 H会場 (10F 会議室1008)

[2-H-1-OP6-6] DPCデータと大阪府の地域がん登録データをリンケージした多施設研究—がん診断時の DPCデータでがん生命予後を予測できるか—

森島 敏隆¹, 松本 吉史¹, 小枝 伸行², 島田 裕子³, 丸濱 勉⁴, 松木 大作⁵, 中田 佳世¹, 宮代 勲¹ (1.大阪国際がんセンター, 2.八尾市立病院, 3.独立行政法人国立病院機構 大阪南医療センター, 4.東住吉森本病院, 5.大阪府済生会吹田病院)

【背景】 多施設の DPCデータを利用する研究では同一人物の複数の病院にまたがったデータを名寄せできないという限界ゆえ、年単位のアウトカム予測モデルを構築することは困難である。そのため、DPCデータから得られる臨床情報が長期予後予測に使えるか否か不明であった。本研究の目的は、DPCデータから得られる臨床情報だけでがん患者の生命予後を予測できるか否かを、長期予後情報を含むデータをリンケージすることによって調べることである。

【方法】 大阪府内の5つのがん診療拠点病院から DPCデータを、予後情報として大阪府の地域がん登録データを手に入して、両データを患者レベルでリンケージした。研究対象者の選択基準を、2010年1月～2012年12月に胃、大腸、肺がんのいずれかで退院した18歳以上の患者とした。除外基準を上皮内がん、またはがん登録で生死を把握できなかった患者とした。Cox比例ハザード回帰モデルを部位別に構築するにあたって、がんの初回入院時の DPCデータの様式1から収集した性、退院時年齢、UICCステージ分類、入院時併存症（Charlson併存症指数に従ってスコア化）、緊急入院の有無、入院時バーセル指数、喫煙指数、Body Mass Indexを説明変数とした。がん登録から収集した全死因死亡を事象、がん診断日から死亡日または最終生存確認日までの日数を生存時間とした。モデルの予後予測の識別力を Harrellの C統計量で評価した。この指標は0.5～1.0の値をとり、0.7～0.8で及第、0.8以上で優れた識別力を有すると判断される。

【結果】 研究対象者は胃、大腸、肺がんの順に（以下、同じ）953、793、760人、死亡者は408、312、498人、生存時間の中央値は1495、1548、602日だった。C統計量は0.84、0.78、0.76だった。

【結論】 どの部位においても予後予測能は及第以上であった。初回入院時の DPCデータから得られる臨床情報だけでもがん患者の予後予測がある程度可能であることがわかった。

DPC データと大阪府の地域がん登録の統合データベース研究

- がん診断時の DPC データによる生命予後の予測能の検証 -

森島 敏隆^{*1}、松本 吉史^{*1}、小枝 伸行^{*2}、島田 裕子^{*3}、
丸濱 勉^{*4}、松木 大作^{*5}、中田 佳世^{*1}、宮代 勲^{*1}

*1 大阪国際がんセンター、*2 八尾市立病院、*3 独立行政法人国立病院機構 大阪南医療センター、
*4 東住吉森本病院、*5 大阪府済生会吹田病院

Record linkage of DPC data and population-based cancer registry data - Predicting survival in stomach, colorectal and lung cancer -

Toshitaka Morishima^{*1}, Yoshifumi Matsumoto^{*1}, Nobuyuki Koeda^{*2}, Hiroko Shimada^{*3},
Tutomu Maruhama^{*4}, Daisaku Matsuki^{*5}, Kayo Nakata^{*1}, Isao Miyashiro^{*1}

*1 Osaka International Cancer Institute, *2 Yao Municipal Hospital,

*3 National Hospital Organization Osaka Minami Medical Center,

*4 Higashisumiyoshi Morimoto Hospital. *5 Saiseikai Suita Hospital

Background: The administrative data produced under the Diagnosis Procedure Combination Per-Diem Payment System (hereafter, referred to as DPC data) do not contain the information of long-term survival outcomes. This limitation makes it difficult to develop prediction models for cancer survival based on DPC data. The aim of the present study was to examine whether clinical information was useful for such prediction, using a record linkage of DPC data and other data.

Methods: We performed a record linkage of DPC data and the population-based cancer registry data of Osaka Prefecture, Japan. Patients diagnosed with gastric, colorectal, or lung cancer who were discharged between 2010 and 2012 at five designated cancer center hospitals that provided DPC data were included. The following variables that were extracted from DPC data were regarded as prognostic factors: sex, age, stage, comorbidity, emergency admission or not, Barthel Index, Brinkman Index, and body mass index. Information regarding vital status and follow-up time was obtained from the registry data. Discriminative ability of predicting overall survival was assessed using Harrell's C-statistic, which was computed from Cox proportional hazards modelling with all-cause mortality as the dependent variable that was developed for each cancer site.

Results: A total of 953, 793, and 760 patients with gastric, colorectal, and lung cancer (hereafter, shown in the same order) were enrolled, respectively. At the time of analysis, 408, 312, 498 patients were confirmed dead with a median duration of follow-up of 1495, 1548, and 602 days, respectively. C-statistic calculated from the prediction model was 0.84, 0.78, and 0.76, respectively.

Conclusions: We found that clinical information extracted from DPC data has an acceptable or good predictive ability for long-term outcomes among cancer patients.

Keywords: administrative data, registry data, medical record linkage, neoplasms, risk adjustment.

1. 緒論

DPC データは診療報酬請求情報のみならず、臨床情報を含んだデータであり、臨床疫学研究に頻用されている。しかし、DPC データ内での個人を識別する ID 番号は病院ごとに独自に付番されたデータ識別番号だけであるがゆえに、多施設のデータを利用する研究で、複数の病院にまたがった同一人物のデータを名寄せできないという限界がある。急性心筋梗塞や肺炎のような、生命アウトカムが1入院エピソードで決することの多い疾病とは違って、がんの生命アウトカムは疾患の診断から死亡するまでに年単位で経過するので、その間に複数の医療機関にかかることも珍しくない。すなわち、DPC データを使ったがんの臨床疫学研究では、年単位の生命予後
を正確には追跡できない。また、生存期間を算出するときの起算日である診断日が把握できないという限界もある。

医療情報データは DPC データを含めて数多く存在するが、近年、複数のデータの患者個人レベルでの連結(リンケージ)の必要性が論じられている。がんのレジストリデータとして、都道府県単位でがん種を問わないすべてのがん患者を対象に

実施されていた地域がん登録(2016 年症例からは全国がん登録に移行)が挙げられる。大阪府の地域がん登録では、がん診断の 10 年後の生死不明割合が 1~2%という精緻さを誇っている。DPC データに地域がん登録データを連結すれば、DPC データでは得られない生命予後情報を付与することができる。実際、米国ではがん登録データである Surveillance Epidemiology and End Results (SEER)と医事管理データの一種である Medicare (高齢者医療保険)のレセプトデータを連結したデータベースを使って、多数の臨床疫学研究が実施されている。¹⁾

2. 目的

DPC データ単独ではその限界ゆえ、アウトカム予測モデルを構築することは困難であった。しかし今後、アウトカムを施設間・地域間で比較しようとするれば、施設・地域により異なる患者構成を調整するための患者背景因子のリスク調整モデルが必須となる。そこで、本研究の目的を、DPC データから得られる臨床情報だけで構築したリスク調整モデルによって、がん患者の生命予後をどの程度予測できるかを調べるこ

ととした。

3. 方法

3.1 データの入手・連結

3.1.1 地域がん登録

大阪府の全医療機関を対象に、大阪府民のがん罹患情報を収集するレジストリであり、全国がん登録制度の前身である。医療機関からの届出情報と大阪府在住者のがん死亡情報と住民基本台帳ネットワーク(または住民票)を情報源としている。大阪府悪性新生物患者登録資料利用に関する取扱要領に則った手続きで入手した。1レコード(=1腫瘍)に、がん診療に関わった病院名、自施設患者 ID、生年月、性別、患者居住地の市区町村、腫瘍の部位・病理診断・臨床進行度・診断年月・治療の概要・生死区分・生存日数等が含まれている。

3.1.2 DPC データ

2016 年度に大阪府がん診療連携協議会(がん診療拠点病院の協議体)のがん登録・情報提供部会の事業として立ち上げた大阪がん診療実態調査に、大阪府内の5つの厚生労働省指定がん診療連携拠点病院または大阪府指定がん診療拠点病院が参加した。この事業において参加病院が DPC データの EF ファイルと様式1ファイルを提供した。

3.1.3 データの受け渡し・両データの連結

大阪がん診療実態調査事業での DPC データの受け渡しからデータベースの連結までの概要を図1に示す。本文中の A)~E)は図中のものと同じである。

- 当センターの事務担当者が地域がん登録サーバーから該当する地域がん登録データを切り出した。各レコードに事業用患者連番を機械的に付した。
- 事務担当者が事業用患者連番と自施設患者 ID の2つのカラムを切り出した対象症例リストを作成した。このリストを当センターから事業参加病院に送付した。
- 地域がん登録データから自施設患者 ID を削除した、研究用地域がん登録データを事務担当者が研究代表者に手渡した。
- 事業参加病院内で参加病院の担当者が3つのデータ、すなわち B) 対象症例リスト、自施設患者 ID と DPC データのデータ識別番号を結びつける対応表、DPC データを、事業参加病院内で突合して、事業への提出用の DPC データを作成した。D) 提出用 DPC データは通常の DPC データとは異なり、次のような加工を施している。
 - データ識別番号を事業用患者連番に置換している。
 - 患者生年月日を削除している。
 - 対象症例ではないレコードを削除している。事業参加病院の担当者が行うこれらの一連の加工作業の負担を軽くするために、DPC データの加工処理用に開発した専用のソフトウェアを事業参加病院に無料配布した。提出用 DPC データを、事業参加病院の担当者が当センターの事務担当者に送付した。事務担当者が研究代表者に手渡した。
- 研究代表者が C) 研究用地域がん登録データと D) 提出用 DPC データを、事業用患者連番をキーにして連結した。95%の患者のレコードが連結に成功した。

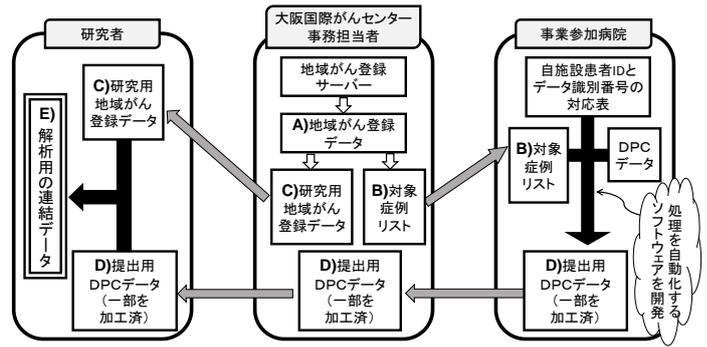


図1 データの受け渡しと連結の流れ

3.1.4 個人情報保護への配慮

「人を対象とする医学系研究に関する倫理指針」に則って実施した。当センターの倫理審査委員会の承認を得た。

本研究で収集する個人に関する情報は、2つのデータの連結に要する最低限のものとし、氏名、患者住所、生年月日の日を収集しない。以下、具体的な配慮を記す。

- ID 番号として自施設患者 ID とデータ識別番号を扱うが、アクセスできるのは事務担当者または事業参加病院の担当者だけである。研究者はアクセスできない。
- 事業参加病院が提供する DPC データでは、①データ識別番号を事業用患者連番に置換、②患者生年月日を削除、③非がん患者のレコードを削除、といった加工を施した。
- 研究者が入手する解析用の連結データに含まれる ID 番号は事業用患者連番だけである。事業用患者連番は、自施設患者 ID やデータ識別番号との間で規則性を有しない。
- 解析用の連結データに含まれる個人に関する情報は、生年月、性別、居住地の市区町村名である。

3.2 データの解析

3.2.1 研究対象患者の選択基準

選択基準を、大阪がん診療実態調査事業に参加した5病院から、胃がん、結腸・直腸がん(以下、大腸がんとする)、肺がんを資源最投入病名とした入院エピソードで、2010年1月~2012年12月に退院した18歳以上の患者とした。除外基準を上皮内癌、またはがん登録データで生死を確認できない患者とした。

3.2.2 連結データベースからのデータの抽出

DPC データとがん登録データを連結したデータから、以下の項目を抽出した。

がん診療のための最初の入院エピソード時の DPC データの様式1から、性、退院時年齢、退院年月日、緊急入院の有無、資源最投入病名、入院時併存症、喫煙指数(ブリンクマン指数:1日の喫煙本数×喫煙年数)、身長、体重、入院時の ADL (activities of daily living) スコア、癌取扱い規約に基づくステージ分類(ステージ)を抽出した。このうち、入院時併存症を Charlson Comorbidity Index (CCI: 転移性腫瘍以外の11疾患群を致死リスクに応じて1~4点で重みづけた点数を個人別に集計)²³⁾²⁴⁾に従って点数化した。Body mass index (BMI: 体重/身長²)を算出した。ADL スコアからパーセル指数(100点満点、ADL 良好で高い点数)を算出した。

地域がん登録データ由来の情報として、生死区分、生存日数を抽出した。

3.2.3 統計解析

全死因死亡をイベントとするCox比例ハザード回帰モデルによる生存時間解析をがん部位別に行った。説明変数を、性(男、女)、年齢(<65、65-69、70-74、75-79、≥80)、緊急入院の有無(あり、なし)、CCI(0、1-2、≥3)、ブリンクマン指数(0、≥1、不明)、BMI(<18.5[るい瘦]、18.5-24.9[標準体重]、≥25.0[肥満]、不明)、バーセル指数(100[完全自立]、61-99[基本動作自立]、≤60[部分～全介助]、不明)、ステージ(I、II、III、IV)とした。

モデルの予後予測の識別力をHarrellのC統計量で評価した。⁵⁾この指標はROC(Receiver Operating Characteristic:受信者動作特性)曲線下面積に相当し、0.5~1.0の値をとり、0.7~0.8で及第、0.8以上で良好な識別力、1.0で完全な識別力を有すると判断される。

SAS 9.4を使ってすべての統計解析を実施した。統計学的な有意水準を両側5%とした。

4. 結果

研究対象者の特性を表1に示す。研究対象者は胃、大腸、肺がんの順に(以下、同じ)953人、793人、760人だった。そのうち死亡者(全体に占める割合)は408人(43%)、312人(39%)、498人(66%)だった。生存時間の中央値は1495日、1548日、602日だった。

表1 研究対象者の特性

		胃 (n=953)	大腸 (n=793)	肺 (n=760)
		人数 (%)	人数 (%)	人数 (%)
性別	男性	660 (69)	421 (53)	521 (69)
	<65	262 (28)	230 (29)	202 (27)
	65-69	176 (19)	146 (18)	158 (21)
年齢	70-74	188 (20)	129 (16)	151 (20)
	75-79	189 (20)	139 (18)	133 (18)
	≥80	138 (15)	149 (19)	116 (15)
	I	397 (42)	169 (21)	123 (16)
ステージ	II	71 (8)	159 (20)	23 (3)
	III	99 (10)	213 (27)	63 (8)
	IV	197 (21)	157 (20)	130 (17)
	不明	189 (20)	95 (12)	421 (55)
CCI	0	755 (79)	635 (80)	544 (72)
	1-2	177 (19)	137 (17)	191 (25)
	≥3	21 (2)	21 (3)	25 (3)
緊急入院	あり	83 (9)	116 (15)	67 (9)
ブリンクマン指数	0	464 (49)	487 (61)	232 (31)
	≥1	438 (46)	269 (34)	499 (66)
	不明	51 (5)	37 (5)	29 (4)
	<18.5	114 (12)	121 (15)	95 (13)
BMI	18.5-24.9	630 (66)	488 (62)	507 (67)
	≥25.0	162 (17)	125 (16)	111 (15)
	不明	47 (5)	59 (7)	47 (6)
バーセル指数	100	834 (88)	657 (83)	632 (83)
	61-99	53 (6)	65 (8)	52 (7)
	0-60	66 (7)	71 (9)	76 (10)
生死区分	死亡	408 (43)	312 (39)	498 (66)
生存日数中央値		1495	1548	602
(四分位範囲)		(434-1702)	(701-1739)	(253-1544)

生存日数以外を、人数(括弧内は全体に占める割合)で表す。

CCI, Charlson Comorbidity Index; BMI, body mass index.

Cox比例ハザード回帰モデルによる生存時間解析の結果を表2に示す。統計学的有意に死亡ハザード比が1以上になった説明変数は、肺がんでの男性; 65歳未満と比較したときの、胃がんと大腸がんでの75歳以上の年齢、肺がんでの70歳以上の年齢; ステージIと比較したときの、胃がんと肺がんでのステージII以上、大腸がんでのステージIII以上; CCI=0と比較したときの、胃がんでのCCI=1-2、大腸がんでのCCI≥3; 胃がんと肺がんでの緊急入院歴; ブリンクマン指数=0と比較したときの、肺がんでのブリンクマン指数≥1; 標準体重と比較したときの、胃がんでのるい瘦; ADLの完全自立と比較したときの、胃がんと肺がんでの基本動作自立、または部分～全介助、大腸がんでの部分～全介助だった。統計学的有意に死亡ハザード比が1未満になった説明変数は、標準体重と比較したときの、肺がんでの肥満だった。

C統計量を表2に示す。胃、大腸、肺がんの順に0.84、0.78、0.76だった。

5. 考察

本研究ではDPCデータの様式1から得られる臨床情報に、がん登録データから得られる生命予後情報を付加することによって、DPCデータだけから得られる臨床情報で胃、大腸、肺がん患者の生命予後をどの程度予測できるかを調べた。

C統計量で評価したモデルの予後予測能では、胃がんが良好、大腸と肺がんが及第だった。すなわち、DPCデータから得られる臨床情報で予後予測がある程度は可能であることが示唆された。

リスク調整モデルの開発は、医療機関別または地域別に生存率を算出して比較するときに、患者集団の背景因子を補正するために必須となるものである。本研究はDPCデータ単独では為し得ない解析を、他のデータの連結という手法を使って可能にした事例である。

6. 結論

DPCデータと地域がん登録データを患者個人レベルで連結したデータを使った分析を行った。

7. 謝辞

DPCデータの抽出・連結・暗号化に貴重な助言・多大な協力をいただいた国立がん研究センターの東尚弘先生にこの場を借りて深く御礼申し上げます。

8. 参考文献

- Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care*. Aug 2002;40(8 Suppl):IV-3-18.
- Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373-383.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. Nov 2005;43:1130-1139.
- Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. Mar 2011;173:676-682.
- Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy,

表2 Cox 比例ハザード回帰モデルによる生存時間解析の結果

		胃		大腸		肺	
		ハザード比 (95%CI)	P 値	ハザード比 (95%CI)	P 値	ハザード比 (95%CI)	P 値
性別(参照:女性)	男性	1.22 (0.96-1.56)	0.109	1.08 (0.83-1.39)	0.59	1.57 (1.21-2.03)	<0.001
	65-69	1.30 (0.95-1.77)	0.102	1.18 (0.82-1.69)	0.38	1.14 (0.86-1.51)	0.35
年齢階級 (参照:<65)	70-74	1.26 (0.92-1.73)	0.153	1.39 (0.94-2.05)	0.100	1.80 (1.35-2.40)	<0.001
	75-79	1.40 (1.02-1.91)	0.035	2.12 (1.49-3.02)	<0.001	2.05 (1.53-2.74)	<0.001
	≥80	2.44 (1.75-3.40)	<0.001	2.62 (1.84-3.73)	<0.001	2.57 (1.90-3.48)	<0.001
ステージ (参照: I)	II	2.61 (1.59-4.29)	<0.001	1.18 (0.68-2.05)	0.56	2.95 (1.44-6.04)	0.003
	III	7.24 (4.95-10.58)	<0.001	3.05 (1.92-4.86)	<0.001	7.47 (4.43-12.59)	<0.001
	IV	27.95 (20.02-39.03)	<0.001	15.04 (9.53-23.75)	<0.001	12.97 (8.05-20.89)	<0.001
	不明	4.04 (2.82-5.79)	<0.001	3.72 (2.21-6.24)	<0.001	5.60 (3.58-8.77)	<0.001
CCI (参照:0)	1-2	1.37 (1.07-1.75)	0.013	1.15 (0.87-1.52)	0.34	1.13 (0.92-1.38)	0.24
	≥3	0.92 (0.51-1.69)	0.80	3.99 (2.32-6.88)	<0.001	1.02 (0.62-1.67)	0.94
緊急入院(参照:なし)	あり	1.56 (1.12-2.17)	0.008	1.29 (0.94-1.79)	0.116	1.78 (1.29-2.46)	<0.001
ブリンクマン指数 (参照:0)	≥1	1.10 (0.87-1.38)	0.44	0.83 (0.63-1.10)	0.200	1.30 (1.01-1.67)	0.045
	不明	1.10 (0.72-1.68)	0.67	0.86 (0.53-1.40)	0.55	1.56 (0.97-2.51)	0.069
BMI (参照:18.5-24.9)	<18.5	1.55 (1.18-2.02)	0.002	1.29 (0.96-1.74)	0.092	1.28 (0.99-1.67)	0.065
	≥25.0	1.04 (0.76-1.41)	0.82	0.70 (0.49-1.01)	0.059	0.63 (0.48-0.84)	0.001
	不明	1.19 (0.77-1.84)	0.43	0.93 (0.59-1.48)	0.77	1.50 (1.05-2.15)	0.026
パーセル指数 (参照:100)	61-99	2.08 (1.43-3.01)	<0.001	1.27 (0.85-1.90)	0.25	1.67 (1.21-2.29)	0.002
	0-60	3.25 (2.21-4.77)	<0.001	1.61 (1.10-2.35)	0.014	2.56 (1.85-3.54)	<0.001
C 統計量		0.84		0.78		0.76	

C 統計量以外を、ハザード比 (括弧内は 95%信頼区間) で表す。

CI, 信頼区間; 参照, 参照カテゴリー; CCI, Charlson Comorbidity Index; BMI, body mass index