



ESRI Discussion Paper Series No.374

医療レセプトを用いた悪性新生物における 3年生存率による医療の質の計測

大里隆也, 桑原進, 菊川康彬

April 2023



内閣府経済社会総合研究所

Economic and Social Research Institute

Cabinet Office

Tokyo, Japan

論文は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI ディスカッション・ペーパー・シリーズは、内閣府経済社会総合研究所の研究者および外部研究者によって行われた研究成果をとりまとめたものです。学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

論文は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

医療レセプトを用いた悪性新生物における3年生存率による医療の質の計測

大里隆也, 桑原進, 菊川康彬¹

要旨

内閣府経済社会総合研究所では、GDP統計の推計の精度向上を図ることを目的に、医療の質の変化を反映した価格の把握手法について研究している。本稿は、「レセプト情報・特定健診等情報データベース」のレセプトデータ（悉皆）を使用し、代表的な医療の質指標である生存率の推計方法を検討したものである。特に正確な測定が進んでいる悪性新生物（がん）のデータを用いて、推計方法の有効性を確認した。データの悉皆性を利用し、受診記録が一定期間ない患者を「見なし死亡」とするアルゴリズムを用いれば、肺癌については一定の精度をもって患者の生存率の経年変化（従って医療の質の変化）を推計できることが確認できたが、比較対象の生存率データとの乖離も見られ、限界があることも判明した。

目次

1.	研究の背景と目的	4
2.	先行研究	4
3.	データと手法	8
	（1）NDB 悉皆データによる生存率	8
	（2）3年生存率の定義	9
	（3）患者名寄せ「疑わしきは同一」	11
	（4）生存時間分析	11
4.	算出結果と考察	12
5.	むすび	23
	参考文献	24

¹大里隆也：内閣府経済社会総合研究所研究協力者、株式会社帝国データバンク

桑原進：内閣府経済社会総合研究所総括政策研究官

菊川康彬：内閣府経済社会総合研究所研究協力者、株式会社帝国データバンク

本稿の作成に当たって、田倉智之教授（東京大学）、伊藤由希子教授（津田塾大学）、原湖楠研究員（アリゾナ大学）より、有益なアドバイスをいただくとともに、内閣府経済社会総合研究所の増島稔所長、林伴子次長、板倉周一郎総括政策研究官、中澤信吾総括政策研究官、出口恭子上席主任研究官、村館靖之研究官、近藤雄介政策調査員から有益なコメントをいただいた。ここに謝意を示したい。なお、本稿における誤り等はすべて著者の責任である。

1. 研究の背景と目的

内閣府経済社会総合研究所では、GDP統計の推計の精度向上を図ることを目的に、医療の質の変化を反映した価格の把握手法について研究してきた。その中で「レセプト情報・特定健診等情報データベース」(以下、NDBという)を用いた分析に取り組んできており、その成果として、西崎他(2022)「医療の質の変化を反映した価格の把握手法に関する研究—レセプトデータ(悉皆)による試算—」を公表したところである。研究を通じてNDBを活用した分析に関するノウハウの蓄積が一定程度進んできたと考えている。

また、医療の質測定の一環として、質を考慮した価格の裏側にある、質的改善を含めた医療の実質アウトプットを測定する方法の研究にも取り組んでいる。医療の質としては、治療の成果(死亡率、平均在院日数、QOLなど)に関連する指標がと取り上げられることが多く、現在、諸外国において指標の選定、傷病別での計測手法等、検討が進められているところである。本稿では、その一つである、生存率に焦点を当て、さらに中でも比較的測定が進んでいる悪性新生物(がん)の生存率について、NDBを用いて算出し活用可能か、検討する。具体的には、国立がんセンターから「3年生存率」が公表されており、これをベンチマークに、NDBを用いて推計した結果と比較し、医療レセプトを用いた医療の質測定の可能性について検討する。

2. 先行研究

本研究には、レセプトを用いた医療の質の測定、生存率の測定、悪性新生物に関する測定という三つの側面があり、それぞれに先行研究が存在する。

(1) 質調整・レセプトに関する先行研究

内閣府経済社会総合研究所でGDP統計の改善のために取り組んできた医療・介護の質測定のこれまでの研究成果がそのまま該当する。藤沢(2012)は、医療・介護質測定の可能性について言及した論文であり、その中でも、がんについては、既存のデータを用いた試算が可能であること、及びその試算値を紹介している。但し、幅広い医療のサービスの中で、試算が可能なのは限られた分野でしかないことは、藤沢(2012)も認めている。

杉原他(2018)は医療・介護の質測定に関する概念整理を行ったものである。この中で、医療・介護の質の調整方法を細分化による質調整と明示的な質調整に分類した。疾病や疾病・治療に着目した分類を可能な限り細分化し、分類内の医療サービスを可能な限り均質にし、質の変化を分類間の移行で捉えるアプローチが細分化による質調整(のちに当研究所では非明示的質調整という表現で統一)で、医療の質を疾病死亡率や再入院率という明示的な指標で捉え、統計的手法を用いてアウトプットに反映させるアプローチが明示的な質調整であるとした。その上で測定対象として実質アウトプットか価格かという分類と組み合わせ、

4種類のアプローチがありうると整理している。

石橋他(2020)は、同じくNDBにおけるサンプリングデータを用いる一方、詳細に分類した傷病別の医療費から価格指数を作成した試みであり、非明示的質調整に取り組んだものである。傷病別ではあるものの、治療が完結していない状態の評価に過ぎないという限界がある。西崎他(2022)は、さらにNDBの特別抽出データを用い、1年間を通じて同一個人について、傷病別の治療費を計算し、価格とした。ただ、レセプトに含まれる診療行為が、同じくレセプトに含まれるどの傷病に対して用いられたかは分からないため、複数の傷病を抱えている患者の場合、どれかに費用を寄せるという妥協が必要になり、価格としての推計精度に限界が生じた。また、このような非明示的な質調整の場合、より高価ではあるがより効果のある治療は、単純に価格の上昇と捉えられるという限界があることも分かった。

今井(2021)は、NDBにおけるサンプリングデータにおいて死亡が確認できたサンプルについて、含まれる診療行為と年次ダミーを説明変数、死亡の有無を被説明変数として推計し、年次ダミー変数の係数の改善から、年々の質的改善を測定しようとした試みである。データはあくまで月次単位でのレセプトのため患者のトレースが困難であり、死亡した個人が死亡にいたるまでに投与された医療サービスをカウントしているわけではなく、限界のある結果となっている。

本研究では、西崎他(2022)で利用したNDBの特別抽出データを用いることによる、同一患者が1年間にどの程度の診療行為を受け、死亡したのか(もしくは生存し続けたのか)を推計し、生存率の改善から医療の質を明示的に測定しようという試みの一つである。

(2) 生存率に関する研究

生存率を用いた分析手法としては、カプラン・マイヤー法、Cox比例ハザード法が医学全般で良く用いられている。

カプラン・マイヤー法による生存率分析は、医学全般、特に悪性新生物に関する分析で多く用いられている。カプラン・マイヤー法自体は、説明変数は経過時間のみであり、生存率(特に実測生存率²⁾の測定手法という側面が強い。カプラン・マイヤー法で生存率を測定する一方、測定された対象が例えば二つのグループに分類されていれば、生存率の違いを統計的に検定することが可能となる。但し、そのためには適切な分類に基づき、生存率が一定以上の精度をもって測定されている必要がある。今回の研究で用いるレセプトは、様々な分類で集計が可能である一方、把握しているデータ、特に死亡については病院で診察を受け、治療を受けている期間に発生したものだけであり、自宅でなくなったりした場合までは、把握できない。

² 死因に関係なく、全ての死亡を計算に含めた生存率で、診断例に対する～年後の生存患者の割合で示される。本稿では、この実測生存率を中心に分析を行うが、実測生存率には他の要因(例えば他の病気の治療における改善)による生存率の改善が含まれてしまうことがあるため、結果の解釈には注意が必要である。

このように今回のレセプトのみを用いた生存率の研究では、院外で亡くなった場合の予後のデータが問題となり、生存率の推計には一定の仮定に基づいた死亡時点の推計などが必要になる。この点で参考となる先行研究として、島田他(2014)がある。島田他は大腸がん検診において重点的に対処すべき年齢を明らかにするために、宮城県がん登録により40歳以上の大腸癌11,415名を把握し、性年齢階級別に検診によるスクリーニング発見癌と外来発見癌の進行度と予後を調査した。予後の判定は、死亡情報も含めてがん登録で把握した情報のみで、生存確認調査は実施していない。その上で、実測生存率では死因が大腸癌以外の場合も死亡例として扱い、補正生存率では死因が大腸癌以外の場合は死亡ではなく観察打ち切りとして計算している。両方の生存率のデータで検診の有無による生存時間の違いについて年齢層別に解析を行い、ログランク検定により、男性55歳から74歳、女性55歳から79歳が大腸がん検診による生存率に有意な違いが出る年齢層であると結論している。

一方、Cox比例ハザードモデルを用いた分析例としては、川辺・藤田(2001)がある。他の手法と比較しつつも、特にCox比例ハザードモデルを用いた分析から、口腔白板症の悪性化要因の抽出を試みている。横浜市立大学医学部口腔外科で診断・経過観察を行った口腔白板症症例を実例として用いているところ、対象症例は、237例、269病変の口腔白板症で、これらはWHOの定義に従った上、非可逆的病変で、組織学的診断の得られた病変である。観察期間は 56.5 ± 43.4 月であり、悪性化はその内12例に各々1病変ずつ、合計12病変が認められていた。説明変数を多数の説明変数からstep wise法で狭めていき、がん履歴の有無、異形成の有無などが影響するという分析結果を報告している。説明変数を絞り込む過程では Kaplan-Meier法により得られた生存率曲線の比較も併用している。

(3) 悪性新生物に関する先行研究

悪性新生物の生存率については、法制度の整備をも伴う形で、データの集積が進んでおり、最も精度の高い生存率の推計が可能となっている。レセプトを用いた生存率の推計と対比し、限界を明示するためにも、説明したい。

宮原(2022)が、説明しているように、がんは国民の多くの死因であり、死亡数、罹患数、生存率などについて把握し、予防や検診に役立てたいというニーズが強く認識され、病院ごと、さらには自治体ごとに独自に発展していた。データを共有し、さらには患者を追跡して分析できるようにすれば、さらに有効な予防戦略、治療戦略が立てられることから、ついには、「がん登録等の推進に関する法律(平成25年12月13日法律第111号)」が成立し、データの管理・利用が可能となった。同法により、病院又は指定された診療所の管理者が原発性のがんについて、当該病院等における初回の診断が行われたとき、一定の期間内に、その診療の過程で得られた当該原発性のがんに関する情報(図表1)を当該病院等の所在地の都道府県知事に届け出ることが義務づけられている(第6条)。

がん登録のデータは、各病院により登録されたデータだけではなく、人口動態統計とも紐づけられており、死亡の確認も行われている点、他の地域に転出した人の追跡も行っている

点で、画期的であり、各種の医療に関するデータの中でもがん統計は最も整備されていると言えよう。

図表1 届け出の対象となっている患者のがんに関する情報³

- (1) がんと診断された人の氏名、性別、生年月日、住所
- (2) がんの診断を行った医療機関名
- (3) がんの診断を受けた日
- (4) がんの種類
- (5) がんの進行度
- (6) がんの発見の経緯
- (7) がんの治療内容
- (8) (死亡した場合は) 死亡日
- (9) その他

なお現在の体制によるデータの蓄積は2016年に始まったばかりであり、当面は、生存率などの推計には、それまで培った、(1) 全国がんセンター協議会の加盟施設のデータ、(2) がん診療連携拠点病院等の院内がん登録のデータ、(3) 都道府県が実施していた地域がん登録のデータに基づく推計値が利用可能となっている。

このうち、最も詳細なものが、国立がん研究センターが公表している「がん診療連携拠点病院等院内がん登録生存率集計報告書」である。3年、5年、10年の生存率の報告書が公表されているが、そのうち5年生存率報告書では、がん診療連携拠点病院等全体での病期、観血的治療の実施別に生存率を推定するだけでなく、都道府県、施設別の生存率集計においても、がんの病期別に生存率を推定している。3年生存率報告書は、5年生存率報告書の作成に必要なデータの蓄積を待たず、公表できる数字をユーザーの利便性のため早めに公表しているものである。90%以上、予後が確認できているデータを持つ機関のデータのみを用いることにより、推計の精度を確報しようとしている。ただし、がん登録が始まる前の院内がん登録情報にのみ依存しているところ、報告書の注意書きでも、がん患者の生存率には、院内がん登録情報としてデータ収集しているがんの病期、年齢、性別だけでなく、対象者の併存疾患の有無や身体機能の状態にも大きく影響を受けることが分かっており、特に施設間では高齢者の多い施設、少ない施設、他の病気をもつ患者の多い施設、少ない施設があるため、単純に生存率を比較して、その施設の治療の善し悪しを論ずることはできない旨、記載されている。本研究で用いるNDB特別抽出データは、保険診療を受けた患者はすべて網羅する悉皆データであり、院内がん登録に参加した病院だけのデータと比較し、患者の偏りが無いという点では、優れている。

³ 国立がん研究センターウェブサイト (<https://ganjoho.jp/public/institution/registry/national.html>) より引用

図表2 国立がん研究センターによる診断年別3年実測生存率(%)

実測	2012	2013	2014	2015
全がん	67.2	67.5	67.6	68.5
50 歳代	77.3	78.1	78.6	79.9
60 歳代	71.3	71.9	72.4	73.5
70 歳代	64.6	65.3	66.0	66.9
80 歳以上	47.3	47.4	47.2	48.4

以上のような限界はあるものの、公表されている3年実測生存率のデータを見ると(図表2)、毎年着実に改善している姿が浮かび上がってきており、がん治療という側面で、医療の質が改善している様子が測定できていると考えられる。医療の質測定が目指す姿の一つが、部分的ではあるものの、実現されつつある。今後は、がん登録制度により、死亡の確認も含め全数ベースでの測定が期待される。

一方、NDB 特別抽出データに含まれるレセプトのみでは、西崎他(2022)にあるように、たとえ地域を超えて紐づけできたとしても、治療中に亡くなり、レセプトに死亡情報が掲載されたケースしか死亡は確認できない。また、主傷病も正確ではない。レセプトの掲載された診療行為ががんに対するものなのかどうかも、必ずしも明確ではない。

3. データと手法

本研究では、先行研究で述べた NDB 活用の限界を考察しながら、生存率を算出するために最大限尤もらしいデータ処理の方法を検討する。ここでは、利用データである NDB に関する整理と3年生存率の定義、患者名寄せ、生存時間分析手法について説明する。

(1) NDB 悉皆データによる生存率

NDB(National Data Base)は、レセプト情報・特定健診等情報データベースの略称であり、レセプト情報と特定健診等情報が格納されている。今回推計に用いるデータは、NDBからの特別抽出である。レセプト(正式名称「診療(調剤)報酬明細書」とは、医療機関等が患者負担額以外の負担分(保険者負担分、公費負担者分、高額療養費(現物高額等)を保険者等に請求する「請求書(診療(調剤)報酬の明細)」である。国民皆保険という日本の保険制度では、患者の医療費は保険機関が負担する仕組みとなっており、保険機関の負担分を後から請求するために、毎月1枚、患者ごとに受診のあった医療機関が発行するものである。医師が患者に提供する医療行為や検査、処方する薬は、医療行為の内容や傷病、医療機関の規模によって点数が決まっており、診療報酬明細書では、患者の傷病に対する医療行為などを記入するだけでなく、それに応じた点数も計算されている。

具体的には、①診療年月日、②都道府県番号・医療機関コード、③公費負担情報、④保険証情報（「保険者番号」「保険の種類」「本人か家族か」「被保険車掌の記号・番号」、⑤患者の氏名・生年月日等、⑥特記事項、⑦医療機関の名称と所在地、⑧傷病名と該当のコード、⑨診察開始日、⑩転帰（傷病に対して診療がどのような結果で終わったかを治癒、中止（転医含む）、死亡の3分類で記載、⑪診療実日数（医療保険と公費で行われた診療日数）⑫診療回数・診療点数・公費分点数、⑬摘要欄が含まれる。

このように、レセプトには診療にかかわる情報が多く含まれ、かつ、NDBでは、すべての患者のデータが含まれるという特徴があり、様々な分析がすぐにでも可能なように思えるが、西崎他(2022)が紹介するように、実際には様々な課題が存在し、患者の追跡、傷病の特定において現時点では、我々は限界があると考えている。

例えば、まず傷病名である。傷病名には主傷病名だけでなく、併存疾患等も記載され、行われた診療行為は別の欄にまとめて記載されている。どちらの治療に費用が掛かったのかは、明確には分離できない。

次に、生存率についてである。レセプトには、治療中に亡くなった場合は、前述したように転記欄の選択肢：「死亡」が選択される。但しこれは、病院内で死亡したケースに限られる。レセプトのみに基づき、患者の死亡を判定すると、自宅療養中の死亡を除外するため、過少推計となる。従って、本研究では、別途、死亡を判定する方法を考えつ必要がある。

また、個人の特定も限界がある。保険証情報だけで特定できそうであるが、実際には多くの人の保険番号が、様々な都合で変わる。同じ会社に勤めていても、転勤等で所属する保険組合が変われば、変わってしまう。戻ってきても違う番号が割り当てられることになる。

（2）3年生存率の定義

研究対象である「がん」患者の特定、「死亡」の確認について説明する。本研究では、2011年4月から2019年3月までのレセプトデータを利用する。傷病については、ICD-10の3桁分類を用い、主傷病が「がん」（ICD-10におけるC00-C97）の患者のみを対象とする。

また、現在入手しているNDBのデータセットが年度を単位にデータが収納されているため、各年4月時点で「がん」を新規発症したケースのみを扱う。ここで、患者が「がん」を新規発症とするために、前年の4月から発症した当年3月までの1年間に、受診した傷病で「がん」が含まれていないことを条件とする。事前の非傷病観測期間によって対象外となる患者数は、2012年4月の医科・DPCレセプトにおけるがん患者数が約15.8万人に対して、約12.7万人であり、がん患者の約77%が対象外となった。

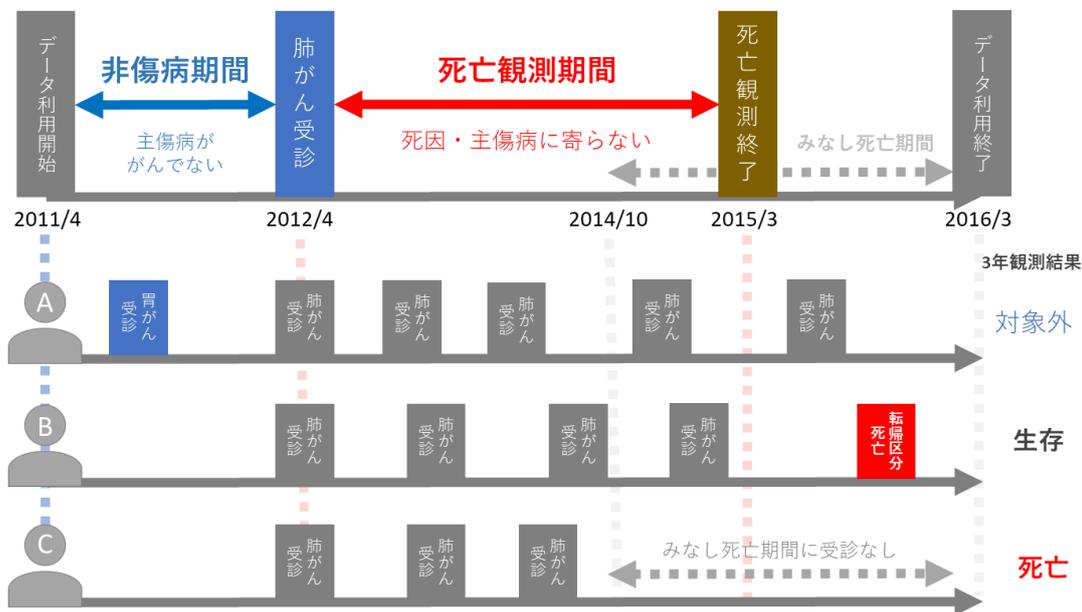
死亡については、病院外で死亡したケースを含めるために、転記区分が死亡である場合に加え、発症から2年7カ月以上4年以内に受診記録がなくなった場合を「死亡」と見な

すこととする。受診期間を1年6カ月としたのは、がんを罹患後は毎年の定期検診を行うことが期待されるが、ある程度のタイムラグを考慮するために半年猶予を持っているためである。ただし、がんの場合、3年再発がなければ完解とされており、本来は3年間の離脱を観測することが望ましい。

データは電子レセプトの集計率が高まる2011年4月以降を利用期間とし、本研究において観測できる時点は、2012年から2015年における各4月時点の4時点分が対象となる。利用可能なデータセットが2019年3月までであるにもかかわらず2015年が最後になるのは、開始時点から見なし死亡期間も含めて4年間のデータが必要となるためである。

図表3では、2012年4月にがんで受診した3人の患者の判別パターンを示している。患者Aは4月の受診以前に「胃がん」で受診した記録があるため、集計の対象外となる。患者Bは、4月に初めて「肺がん」で受診したため集計対象であり、観測期間である3年経過（2015年3月）以降に死亡しているため、本集計では生存となる。一方、患者Cは、集計対象であるが、2014年10月から2016年3月にかけて受信記録がないことから、死亡したと判別する。

図表3 本研究における生存・死亡パターン：2012年4月発症患者を対象とした判別例



(3) 患者名寄せ「疑わしきは同一」

次に、「個人」を特定する方法を説明する。NDBでは、レセプトに含まれた情報を加工し、二つの個人識別番号（以下ID）を各レセプトデータに割り振っている。ID1は「保険番号+被保険者の記号・番号+生年月日+性別」から生成され、ID2は「氏名+生年月日+性別」で生成されている。前述した通り、保険番号は変わり、名前も変わるため、ID1、ID2それぞれでは、個人を特定出来ない。このため、パネルデータとしてNDBを利用する研究者は、二つのIDを組み合わせた名寄せのロジックの開発を進めている。

本研究において、患者が紐づかないことで死亡数観測を過小評価してしまうことを避けるために、西崎(2022)における名寄せロジックを踏襲する。ただし、西崎でのロジックは接続を年単位で行っていたが、本研究では死亡までの長期的な患者の接続が必要となるため、2011年4月から2019年3月までのレセプト利用期間すべてを接続の対象とする。

方法は、まず、2011年4月のデータからID1に対して翌月のレセプトを接続する。ただし、転記区分に死亡がある場合にはその後はデータを接続しない。次に、ID1で接続できなかったレコードをID2で接続する。これを、2019年3月まで実施し、2011年4月のID1と2019年3月時点もしくは最終受診月におけるID1が一致するものを確認した。ID2で名寄せした結果、2019年3月時点もしくは最終受診月におけるID1が複数ある場合でもID2で一人であれば一人とするとして、可能性のあるIDは同一患者として見なしている。

なお、久保(2017)では、「疑わしきは同一でない」という名寄せを実施しており、約5.2%、約600万人の統合が実施されたとしているが、本研究で採用した「疑わしきは同一」では約8.3%、約1,000万人の統合が実施されている。名寄せの方法によって、患者の数が400万人も異なることから、名寄せによるバイアスの影響は小さいとは言えない。

(4) 生存時間分析

以上の3年生存率の定義、個人の特定期間を用いた上で、生存時間分析を行う。

2012年、2013年、2014年、2015年の各4月において新規発症した患者が、それぞれ3年間の間に、どの程度死亡したのかを算出し、年ごとの改善度合いを検定することにより医療の質の改善を推定する。より具体的には開始年ごとにカプラン・マイヤー法によりハザード率を算出し、対象年の際をログランク検定、及びログランク検定とほぼ同等であるがデータ数の多い前の期のデータを多く使う一般化Wilcoxon検定を行う。

なお、転帰区分が死亡の場合には転帰区分が記載された月を死亡時点とするが、死亡見なしの場合の死亡時点は最終受診月の翌月とする。生存患者の場合には、最終の受診月を打ち切り月とする。

4. 算出結果と考察

本章では、悉皆データを用いた集計・分析結果を報告し、肺がんとすい臓がんにおける考察を行う。

(1) 推計がん患者数, がんによる死亡者数

図表4は、各年の新規発症患者における、NDBから算出した患者数と死亡者数、見なし死亡者数である。新規発症時の診断群分類がDPCレセプトあるいはDPC・医科レセプト両方を区別して算出している⁴。本稿では「がん」全体、すなわち「悪性新生物」全体(ICD-10の3桁分類でC00-C97)と、推計期間中に有効な新薬が登場した「肺がん」(同C33,C34)、死亡率の高い「膵臓がん」(同C25)について推計を行った。

死亡者数と見なし死亡者数を比較すると、DPC・医科において、見なし死亡者数が多いことが分かる。また、DPCのみと医科も含む場合では、医科を含む場合の方が患者数に対する死亡者数の割合が低く、医科では院内で死亡するケースが少ないというバイアスが存在することが分かる。このことは特に膵臓がんで顕著であり、DPCでは死亡者数は患者数の半数に及ぶ一方、医科を合わせると死亡者数は患者数の1割程度となる。医科データを含む推計には死亡率に下方バイアス(生存率に上方バイアス)がかかると予想される。

図表4 各傷病での対象患者数、死亡者数、見なし死亡者数

傷病	発症年月	DPC データのみ			DPC・医科データ		
		患者数	死亡者数	見なし死亡者数	患者数	死亡者数	見なし死亡者数
悪性新生物	2012年4月	22,261	4,558	5,706	360,787	21,087	87,548
	2013年4月	19,443	4,532	4,069	339,327	22,345	64,144
	2014年4月	18,814	4,546	3,902	341,347	22,895	63,585
	2015年4月	19,256	4,696	3,905	349,077	23,054	62,505
肺がん	2012年4月	2,087	690	615	38,902	3,691	8,694
	2013年4月	1,760	609	454	38,370	3,781	7,318
	2014年4月	1,710	620	421	38,204	3,900	7,241
	2015年4月	1,783	630	402	39,210	3,927	7,103
膵臓がん	2012年4月	815	434	241	14,619	1,463	2,966
	2013年4月	781	431	210	14,927	1,566	2,614
	2014年4月	833	447	260	15,751	1,664	2,704
	2015年4月	816	456	211	16,520	1,721	2,689

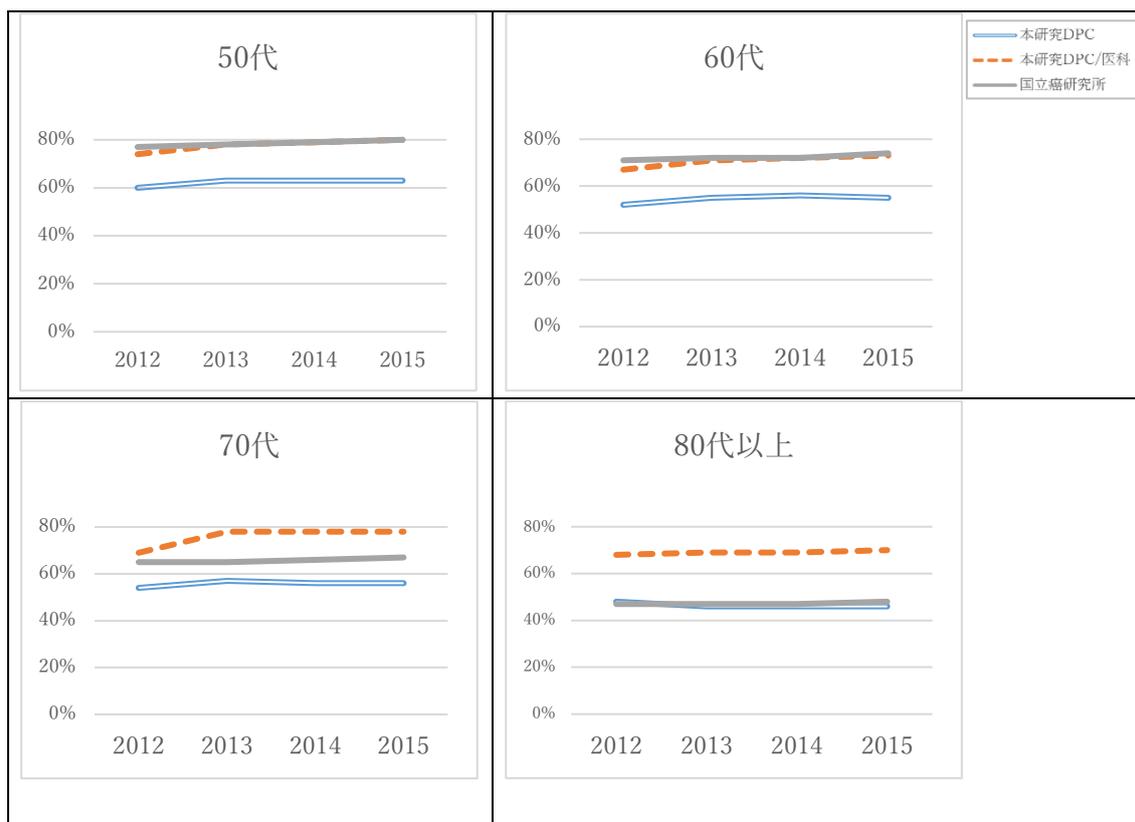
⁴ DPC制度における診療報酬制度については、石橋他(2020)の2.3.2を参照ありたい。また、DPC制度が医療の質に与える影響については、岡崎他(2023)で取り上げている。DPC制度への参加の有無がレセプトデータに与える影響は大きいと考えられる。

(2) 全がん

全がんで生存率を各年4月に発症した患者ごとに計算した生存率の推計結果を年代別にグラフにしたのが図表5である。比較のため国立がん研究センターのデータ(図表2)も掲載している。50代においては、医科を含むデータによる推計値(点線)と国立がん研究センターの値がおおむね一致している。一方、80代以上ではDPCのみのデータによる推計値(二重線)が国立がん研究センターの値と一致していることが読み取れる。

国立がん研究センターでは、全がんの集計対象として「本集計では、原則として新生物<腫瘍>の性状コード3の「悪性、原発部位(悪性新生物<腫瘍>)」の例を集計対象とした。但し、脳・中枢神経系に発生した腫瘍性疾患については、良性、良性又は悪性の別不詳の例を含めて集計対象とした。また、GISTの良性または悪性の別不詳を集計対象に含めた。」とあり、ICD-10とは別の基準で分類されているため、そもそも完全には対象が一致していないと思われる。いずれにせよ、年代により乖離するデータが異なる理由は現時点では不明である⁵。

図表5 年代別「がん」生存率の推移



⁵ 先行研究の部分で指摘したように、国立がん研究センター3年生存率に用いられたデータはがん診療連携拠点病院から得られたものに限定されている。DPCデータもDPC参加病院からのデータに限定され、一定の偏りが存在する。施設特性などをコントロールした上で非劣勢試験を行うべきであるが、現時点では取り組めていない。

また、医科を含めた方が全体に生存率が高いことと、大まかな傾向として年がたつにつれ生存率が上昇していることが分かる。80代以上では国立がん研究センターのデータでも改善ははっきりと見えず、DPCデータでの推計結果ではむしろ低下している⁶。

図表6は各年4月発症者の年代別がん生存率の推移の表である。他のデータではおおむね年がたつにつれ生存率が改善しているところ、DPCデータのみの80代以上のデータは2012年の3年生存率が48%のところ、46%に低下している。医科を含む場合、68%から70%に改善している。

次に年次別の生存率の変化を検定する。DPC分類での2012年4月発症ケースと2015年発症ケースでのカプラン・マイヤー法(=積極限法)による年代別(60代、70代、80代以上)生存率の推移について、ログランク検定、一般化Wilcoxon検定を行う。図表7では検定結果を、図表8では生存時間曲線を示す。60代、70代では、赤の2015年の値が青の値を超えている一方、80代以上では関係が逆転している。検定の結果、生存率に違いがあることは確認できるが、80代以上では生存率が悪化したという結果になっている。

図表6 各年4月発症者の年代別がん生存率の推移

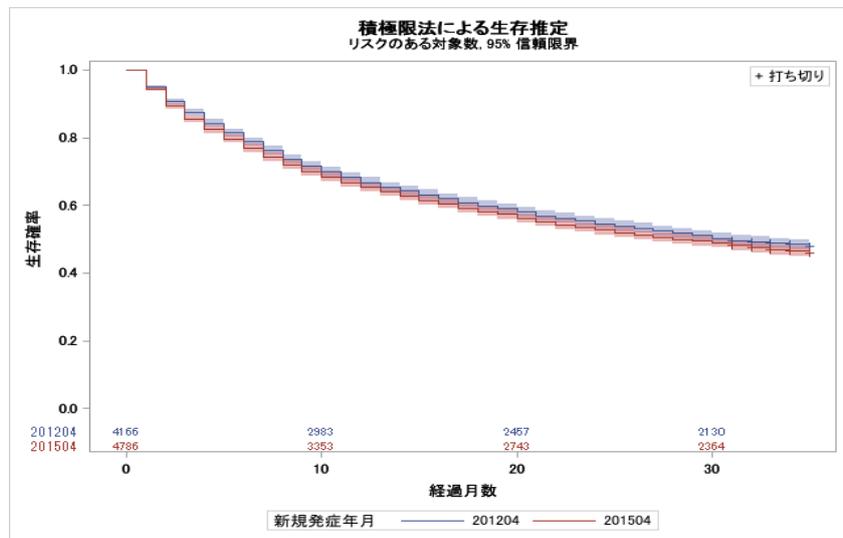
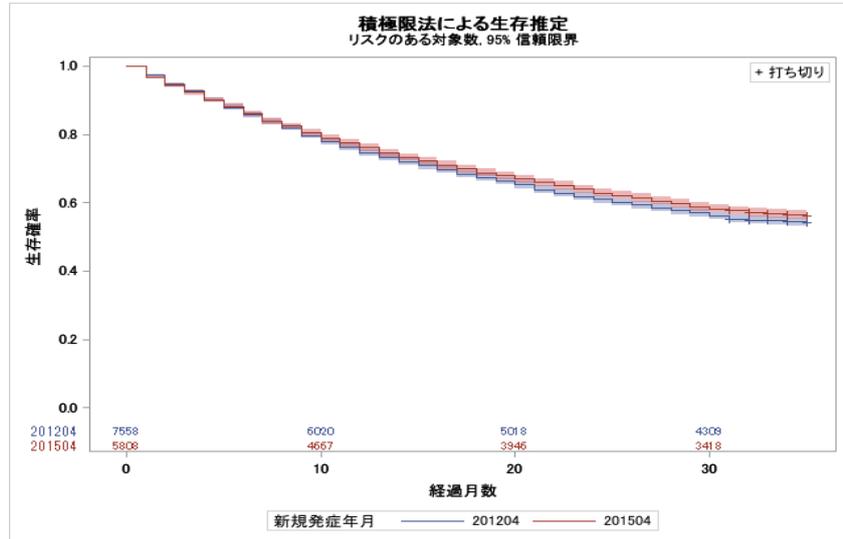
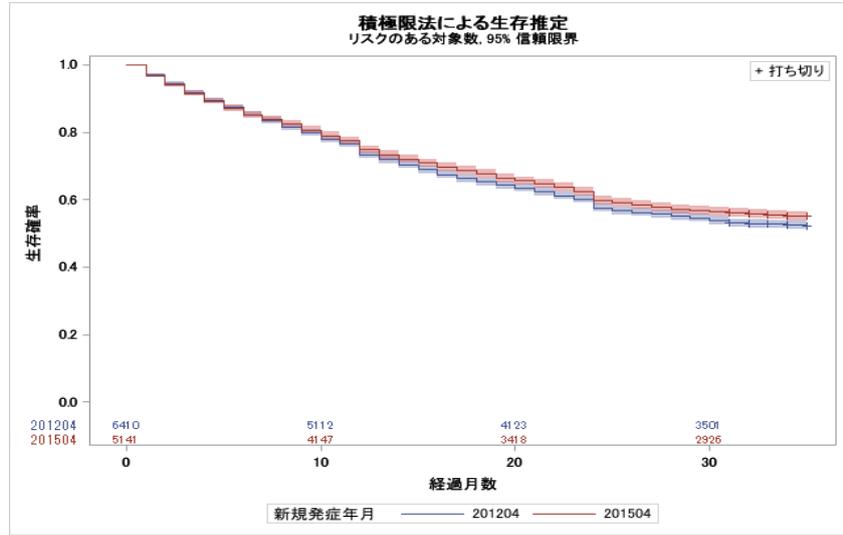
初診年月	生存年数	DPC データのみ				医科・DPC データ			
		50歳代	60歳代	70歳代	80歳以上	50歳代	60歳代	70歳代	80歳以上
2012年4月	1年	81%	76%	76%	68%	89%	86%	86%	84%
	2年	68%	60%	62%	55%	80%	74%	75%	75%
	3年	60%	52%	54%	48%	74%	67%	69%	68%
2013年4月	1年	82%	78%	78%	68%	91%	88%	90%	84%
	2年	70%	63%	65%	54%	83%	77%	83%	75%
	3年	63%	55%	57%	46%	78%	71%	78%	69%
2014年4月	1年	82%	78%	77%	66%	91%	88%	90%	84%
	2年	69%	63%	63%	53%	83%	78%	83%	76%
	3年	63%	56%	56%	46%	79%	72%	78%	69%
2015年4月	1年	81%	77%	77%	67%	91%	88%	91%	85%
	2年	69%	62%	64%	53%	84%	78%	83%	76%
	3年	63%	55%	56%	46%	80%	73%	78%	70%

図表7 各年代の生存率に対する検定結果

	60歳代		70歳代		80歳以上	
	χ^2 乗	p 値	χ^2 乗	p 値	χ^2 乗	p 値
ログランク	8.075	0.005	5.226	0.022	4.090	0.043
Wilcoxon	6.442	0.011	4.476	0.034	4.710	0.030

⁶ 80代以上とオープンエンドで生存率を考えた場合、患者の高齢化による年齢構成比の影響も含まれてしまい、医療の質の改善にも関わらず生存率が下がる可能性がある。

図表8 60代の生存率(上), 70代の生存率(中), 80代以上の生存率(下)の推移



(3) 肺がん

肺がんは、堀池(2021)でこれまでの肺がんに対する薬物療法の変遷をまとめており、免疫チェックポイント阻害薬による薬物療法は、2015年にニボルマブ、2016年にペムブロリズマブが承認され、有意に生存を延長すると報告されている。本研究で用いているデータによって、2012年4月・2015年4月それぞれで新規発症した患者の生存率を比較することができ、薬の改良に関する分析を行うことができる。

同様の生存率推計結果の確認と検定を肺がんについて行くと、以下の図表9、10、11の通りとなる。全体に生存率の改善がみられるが、検定による有意となったのはログランク検定における60代、及び70代である。がん全体ではほとんどの年代で有意となったが、肺がんのみの場合、サンプル数の減少により検出力が低下、有意となりにくい状況となっている。そのような状況下でも、60代及び70代における生存率に有意な差があったことは、薬の投与による影響が考えられる。

なお、肺がんの新薬投与は高額であることから、点数が受診年によって変化することが考えられるため、費用面の変化を明示的に確認する。それぞれの新規発症時期に対する患者の点数を観測期間である3年間の合計を集計し分布を図表12、その分布の代表値を図表13に示す。いずれの年代でも2012年から2015年にかけて最大点数が大きくなっていき、また、外れ値となる点数が増加していることが確認でき、平均値は大きく増加していることが分かる。ただし、60代の場合、中央値などの分位点では発症時期によっても大きな変動は見られず、点数が増加している患者が一部の患者であることがみられた。70代においては、分位点でも増加が確認されており、マン・ホイットニーのU検定でp値0.0164と有意な差があった⁷。

以上のことから、点数が一部患者あるいは全体の患者で増加し、かつ、生存率が向上していることが確認でき、技術進歩の反映がDPCレセプトを用いた分析からも把握することができた。

⁷ 費用（診療報酬点数）と便益（生存率）が計算できるため、費用対効果分析も可能かもしれないが、費用の増大には延命自体がもたらす医療費の増大分も含まれることになるなど、費用と便益に含める範囲について検討すべき課題が多い。より詳細な検討が必要であり、本稿の分析ではそこまで踏み込めない。

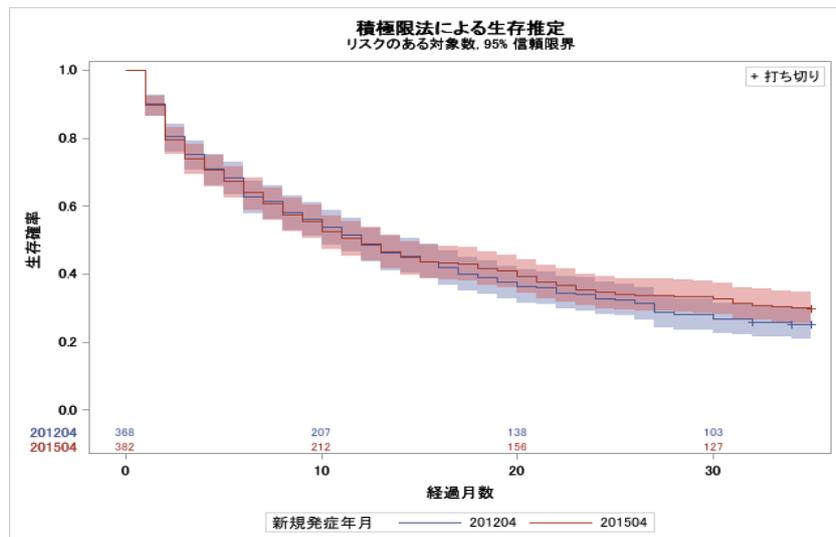
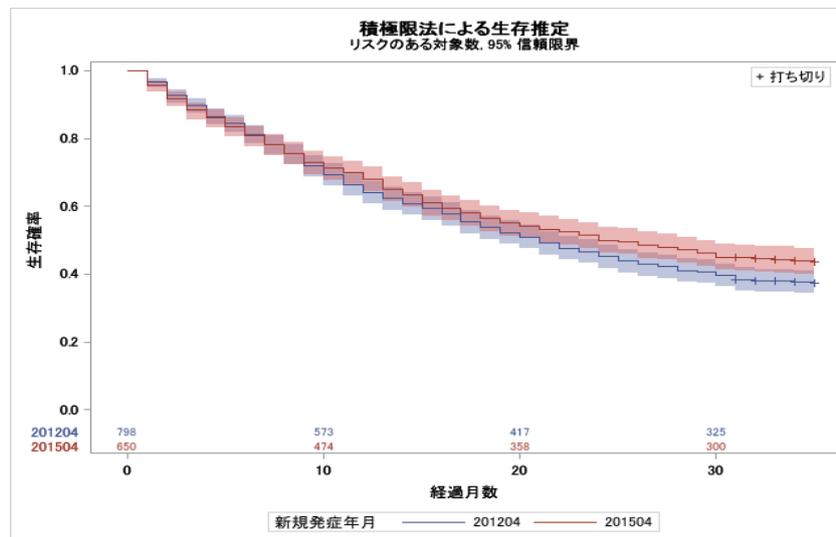
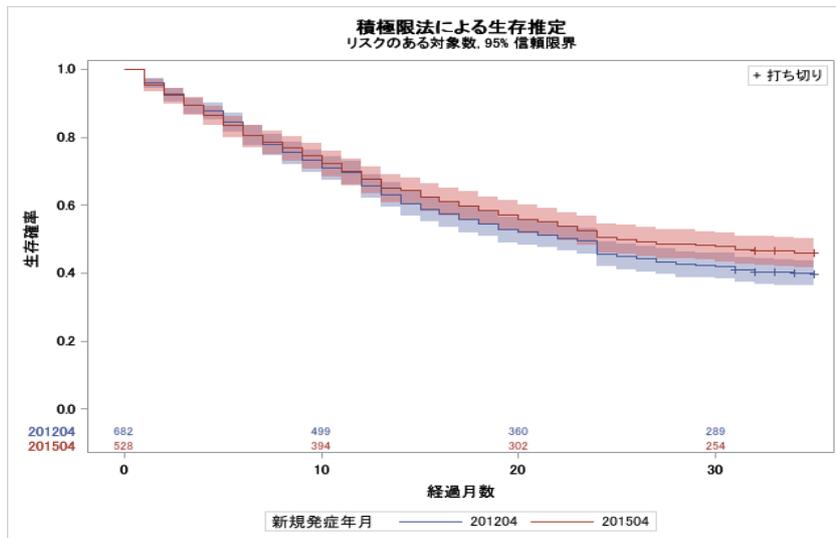
図表9 各年4月発症者の年代別肺がん生存率の推移

初診年月	生存年数	DPC データのみ				医科・DPC データ			
		50 歳代	60 歳代	70 歳代	80 歳以上	50 歳代	60 歳代	70 歳代	80 歳以上
2012 年 4 月	1 年	77%	70%	66%	52%	89%	85%	85%	79%
	2 年	60%	49%	47%	34%	80%	73%	74%	69%
	3 年	49%	40%	37%	25%	74%	67%	68%	62%
2013 年 4 月	1 年	74%	69%	69%	53%	90%	86%	88%	80%
	2 年	57%	51%	50%	36%	82%	75%	80%	69%
	3 年	48%	40%	42%	31%	77%	69%	74%	62%
2014 年 4 月	1 年	67%	71%	67%	54%	90%	85%	88%	80%
	2 年	49%	51%	48%	37%	83%	75%	78%	69%
	3 年	42%	42%	40%	31%	78%	69%	73%	62%
2015 年 4 月	1 年	74%	70%	70%	51%	90%	87%	88%	81%
	2 年	53%	53%	51%	35%	84%	76%	79%	71%
	3 年	46%	46%	44%	30%	79%	71%	73%	64%

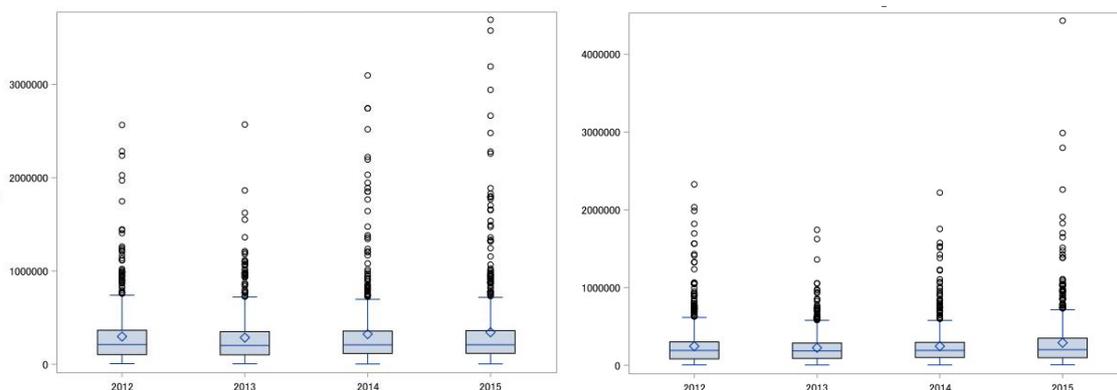
図表10 肺がんにおける検定結果

	60 歳代		70 歳代		80 歳以上	
	χ^2 乗	p 値	χ^2 乗	p 値	χ^2 乗	p 値
ログランク	3.451	0.063	4.136	0.042	0.846	0.358
Wilcoxon	2.082	0.149	2.371	0.124	0.169	0.681

図表 11 肺がんにおける生存率推移：60代(上)，70代(中)，80代以上(下)



図表 12 肺がん新規発症患者の3年間の点数合計の箱ひげ図：60代（左），70代（右）



図表 13 肺がん新規発症患者の3年間の点数合計の代表値：60代（上），70代（下）

60代	2012年4月	2013年4月	2014年4月	2015年4月
平均	296,558	285,209	322,591	342,006
標準偏差	310,966	286,181	394,638	456,593
25%点	103,084	100,101	115,527	116,598
中央値	210,519	201,099	207,497	208,364
75%点	364,009	349,885	356,349	360,320

70代	2012年4月	2013年4月	2014年4月	2015年4月
平均	244,952	224,908	245,108	289,722
標準偏差	262,421	202,018	246,150	357,715
25%点	82,764	88,142	99,410	96,658
中央値	190,339	184,698	190,769	200,562
75%点	299,878	285,386	294,937	349,283

(4) すい臓がん

すい臓がんは、悪性新生物の中でも特に生存率が低く生存率の変化が大きいことから、着目する一つの傷病とした。

図14において、DPCのみ、医科+DPCの生存率の推移を比較している。医科+DPCにおいては、80歳以上でも3年生存率が70%弱と非常に高い水準になっており、国立がん研究センター公表3年生存率が相対で9.6%であったことと比較すると、実態と大きな乖離が存在していることが確認できる。見なし死亡患者数が転帰区分で確認された死亡患者の2倍程度存在しており、医科レセプトデータ自体に大きなバイアスが存在していると考えられる。医科レセプトの活用はデータバイアスを許容しながら分析を行う必要がある。

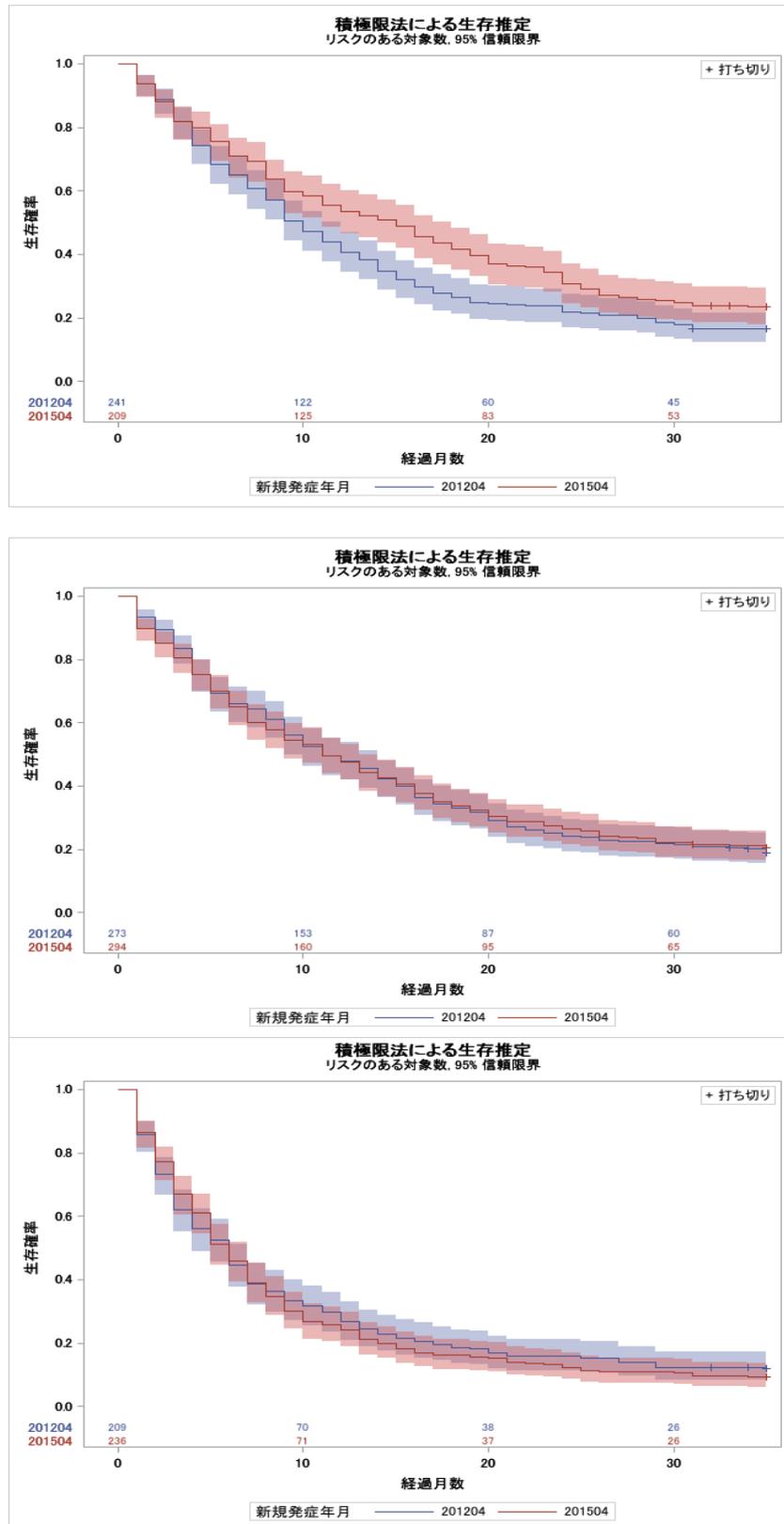
すい臓がんにおいてもDPCデータを用いた同様の分析を行い、図表15・16にて生存時間解析の結果を示す。60代のみで有意に改善していることが確認できたが、サンプル数が少なく、生存時間曲線が不安定な形状となっている。

すい臓がんでの生存時間曲線を安定させるため、サンプル数を増加させる試みとして、4月単月の患者を対象としたのを、4月～翌年3月までの年度単位での新規発症患者を集計し、同様の分析を行った。60代・70代での結果を図表17に示す。比較時期が、データで取り得る最大の差として2012年度（2012年4月～2013年3月での新規発症患者）と2014年度（2014年4月～2015年3月での新規発症患者）の比較となるが、いずれの年代においても有意な差は確認できなかった。

図表14 各年4月発症者の年代別はいん臓がん生存率の推移

初診年月	生存年数	DPC データのみ				医科・DPC データ			
		50歳代	60歳代	70歳代	80歳以上	50歳代	60歳代	70歳代	80歳以上
2012年4月	1年	54%	44%	49%	30%	88%	83%	84%	80%
	2年	30%	24%	25%	16%	81%	72%	74%	73%
	3年	23%	17%	19%	12%	77%	67%	69%	67%
2013年4月	1年	65%	42%	51%	29%	90%	84%	87%	79%
	2年	42%	22%	29%	12%	82%	73%	79%	71%
	3年	38%	16%	21%	8%	77%	69%	74%	66%
2014年4月	1年	51%	46%	43%	25%	90%	85%	86%	79%
	2年	26%	25%	25%	13%	83%	74%	78%	72%
	3年	23%	17%	15%	9%	79%	69%	74%	67%
2015年4月	1年	52%	56%	50%	26%	91%	86%	87%	80%
	2年	31%	34%	28%	13%	83%	76%	78%	72%
	3年	19%	23%	20%	9%	80%	71%	73%	68%

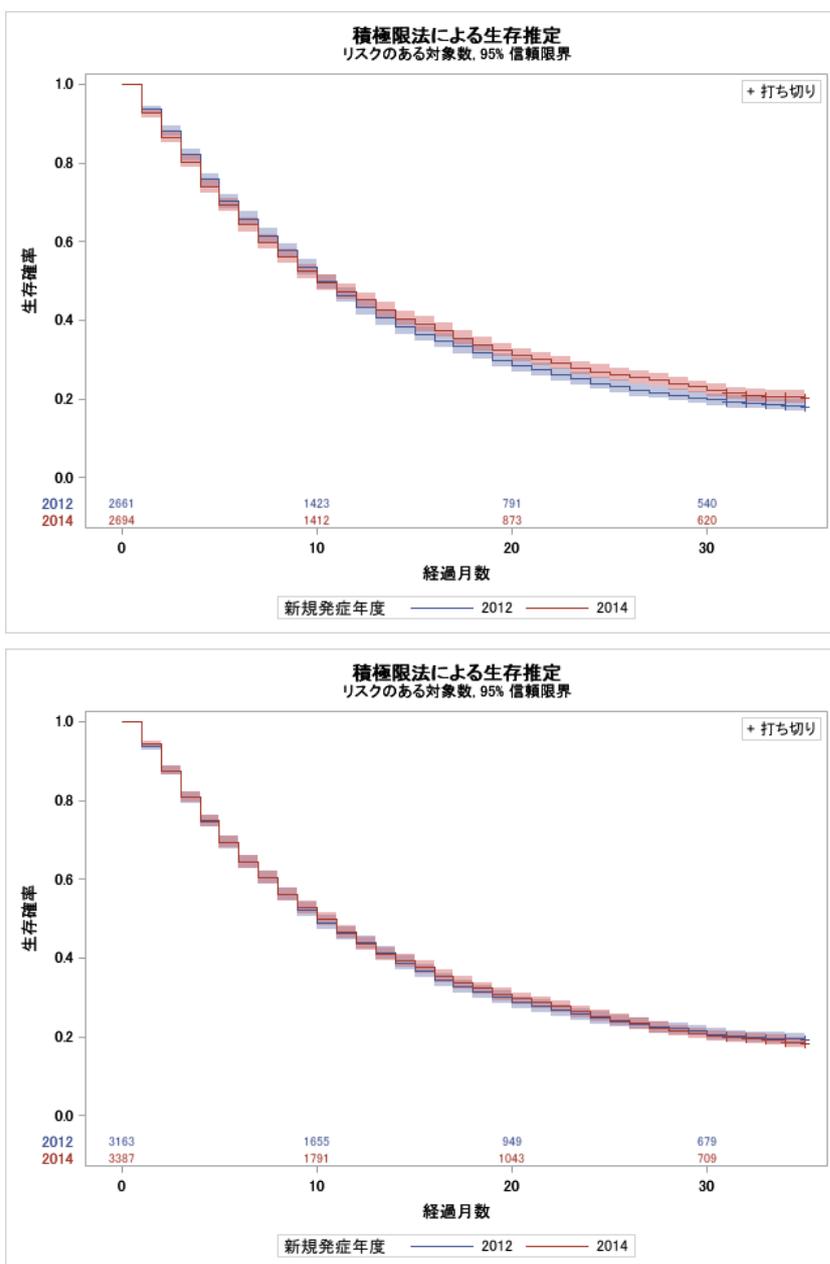
図表 15 すい臓がんにおける生存率推移：60代(上)，70代(中)，80代以上(下)



図表 16 すい臓がんにおける検定結果

	60 歳代		70 歳代		80 歳以上	
	χ^2 乗	p 値	χ^2 乗	p 値	χ^2 乗	p 値
ログランク	6.296	0.012	0.020	0.888	0.262	0.609
Wilcoxon	5.937	0.015	0.047	0.829	0.001	0.912

図表 17 年度単位でのすい臓がん生存率推移：60 代（上）、70 代（下）



5. むすび

本研究では、医療の質を示す代表的な指標である「生存率」に着目し、生存率が計算しやすく、かつ、統計が明確な比較対象データの存在する「がん」についてNDBを利用して推計を行った。これにより、本研究で用いたNDBや手法の有効性を判断しようという試みである。データの質に差が存在すると考えられるDPCレセプトと医科を含めた2パターンで推計した。

肺がんでは、医療点数の増加と生存率が改善した年代が確認でき、先行研究での技術進歩の反映がDPCレセプトデータからも得られる結果となった。すい臓がんでは、DPCレセプトと医科レセプトの両方を用いた生存率では、公表されている統計との乖離が大きいことが確認できた。

国立がん研究センターによる「がん診療連携拠点病院等院内がん登録生存率集計報告書」と比較し、年代によって一致する結果が異なり、DPCと医科レセプトでは、死亡の確認に差があることが分かる。これらの差は、初診における重症度がコントロールされていない、セレクションバイアスによるところが大きいと考えられるが、どこに原因があるのか、本研究では特定できていない。

これらの結果から、NDBによって生存率の改善を検定することができるなど、部分的であっても医療の質を明示的に測定するデフレータの作成につながる成果があったと考える。

今後の課題としては、見なし死亡におけるバイアスの除去が依然必要である。本研究の集計では、医科データを含めると見なし死亡数が転帰区分による死亡数の3倍以上となっている。見なし死亡を精緻化するためには、調剤レセプトや歯科レセプトでの受診記録を利用し、見なし死亡期間での患者の観測率を向上させることや、見なし死亡期間の観測をより長期にすることなどが考えられる。また、患者の名寄せを行う際に、本研究では可能性のあるIDは同一患者としているが、関係ない多くの人が紐づけられてしまい、余剰な見なし死亡を招いてしまっている可能性もある。

このようにNDBのみによる推計には限界があることも本研究を通じて理解が進んだ。NDBのみを用いた名寄せによる誤差は正解データが存在しえないため、あくまで統合された数字などを比較するしか改善の方法はないが、患者情報にマイナンバーが紐づけられることで、名寄せバイアスを除くことができるため、マイナンバー制度における今後の発展に期待したい。

本研究では、悪性新生物との比較に留まっているが、NDBの活用でこれまでできていない観測も実施することができるようになる。悪性新生物以外の傷病の生存率の算出、特に、希少疾病は悉皆データでしかできない対象である。また、悪性新生物についても、がん登録以前の患者の状態の把握や悪性新生物以外の合併症も含めた分析が可能となる。医学系における研究ではレセプトに記載された介入の違いから傷病の種類、重症度を判断す

る研究も進んでおり、カルテとの接続がなくてもできることは増えている。医学系の知見が加わればさらなる研究ができよう。さらに言えば医学系における研究の進展を待つことも選択肢の一つであろう。本研究を通じて課題が多いことも認識したが、それにはビッグデータを扱うための計算機を含む相当の研究資源が必要であることも含まれる。

今後は、データ利用における制度の進展を待ちつつ、幅広い知見を活用し、効率よく研究を行うことが望まれる。

参考文献

石橋尚人, 丸山雅章, 桑原進, 石井達也, 川崎暁, 西崎寿美, 村館靖之, 大里隆也, 菊川康彬 (2020) 「医療の質の変化を反映した価格の把握手法に関する研究－推計法の検討とレセプトデータによる試算－」, ESRI Research Note No.56, 内閣府経済社会総合研究所

今井健太郎, 大里隆也, 菊川康彬, 村館靖之, 石橋尚人(2021) 「医療の質の変化を反映した価格の把握手法に関する研究－死亡関数を用いた医療の質の統計学的計測－」 ESRI Research Note No.58, 内閣府経済社会総合研究所

岡崎康平, 杉原茂, 田倉智之, 村館靖之 (2020) 「医療費の包括支払い制度が医療の質に及ぼす影響－マッチング推定による検証－」, ESRI Research Note No.74, 内閣府経済社会総合研究所

川辺良一, 藤田浄秀(2001) 「生存時間分析: 口腔白板症の悪性化に関する研究への応用」, 日口粘膜誌第7巻第2号, 2001年12月

久保慎一郎, 野田龍也, 西岡祐一, 明神大也, 東野恒之, 松居宏樹, 加藤源太, 今村知明(2017) 「レセプト情報・特定健診等情報データベース (NDB) 利用促進に向けた取り組み - 患者突合 (名寄せ) の手法開発と検証 - 」 奈良県立医科大学

島田剛延, 千葉隆士, 加藤勝章, 渋谷大助(2014) 「大腸がん検診において重点的に対処すべき年齢層－生存率と期待生存年数からの考察－」, 日本消化器がん検診学会雑誌, Vol.52(5), 2014年9月

杉原茂, 市川恭子, 今井健太郎, 野口良平, 岡崎康平, 小池健太(2018) 「医療の質の変化を反映した実質アウトプット・価格の把握～方法論の整理～」, ESRI Research Note No.36, 内閣府経済社会総合研究所

西崎寿美, 近藤雄介, 大里隆也, 菊川康彬(2022)「医療の質の変化を反映した価格の把握手法に関する研究—レセプトデータ(悉皆)による試算—」, ESRI Research Note No.65, 内閣府経済社会総合研究所

野口良平, 市川恭子, 藤森裕美, 岡崎康平, 小池健太, 石橋尚人(2019)「諸外国における医療・介護の質の変化を反映した価格・実質アウトプットの把握手法～各国ヒアリングの結果～」, ESRI Research Note No.48, 内閣府経済社会総合研究所

藤澤美恵子(2012)「国民経済計算における医療のアウトプット計測についての考察」, 季刊国民経済計算, No.149, 2012年9月, 内閣府経済社会総合研究所国民経済計算部

堀池篤(2021)「肺がんに対する薬物療法の最前線」, 昭和学会誌, 第81巻, 第3号

宮原裕(2022)「我が国におけるがん罹患の実態—がん登録の歴史, 生存率—」, 安田女子大学紀要 50, 2022年