# レセプト情報・特定健診等情報データベース（ＮＤＢ）を

# 利用した研究について

**健康・医療研究会　櫛引 亮**

司会　時間となりましたので、セッションＧ、「レセプト情報・特定健診等情報データベース（ＮＤＢ）を利用した研究について」を開始します。発表者は、健康・医療研究会座長でいらっしゃる、櫛引様でございます。それでは、よろしくお願いいたします。

櫛引　どうもありがとうございます。皆さん、おはようございます。

会場　おはようございます。

櫛引　本日は、健康・医療研究会において、ＮＤＢの第三者提供を受けて研究しようという取り組みについて、進捗状況も含めて、お話ししたいと思います。

概要にも書きましたけれども、ＮＤＢとは、国がレセプトデータ、特定健診および特定保健指導のデータを集めて、データベース化したものです。これを、第三者提供という形で、研究者に対して提供して、分析してもらおうという取り組みが行われています。今回の健康・医療研究会取り組みは、それにアクセスして研究を行おうという目標のプロジェクトでございます。

## 概要

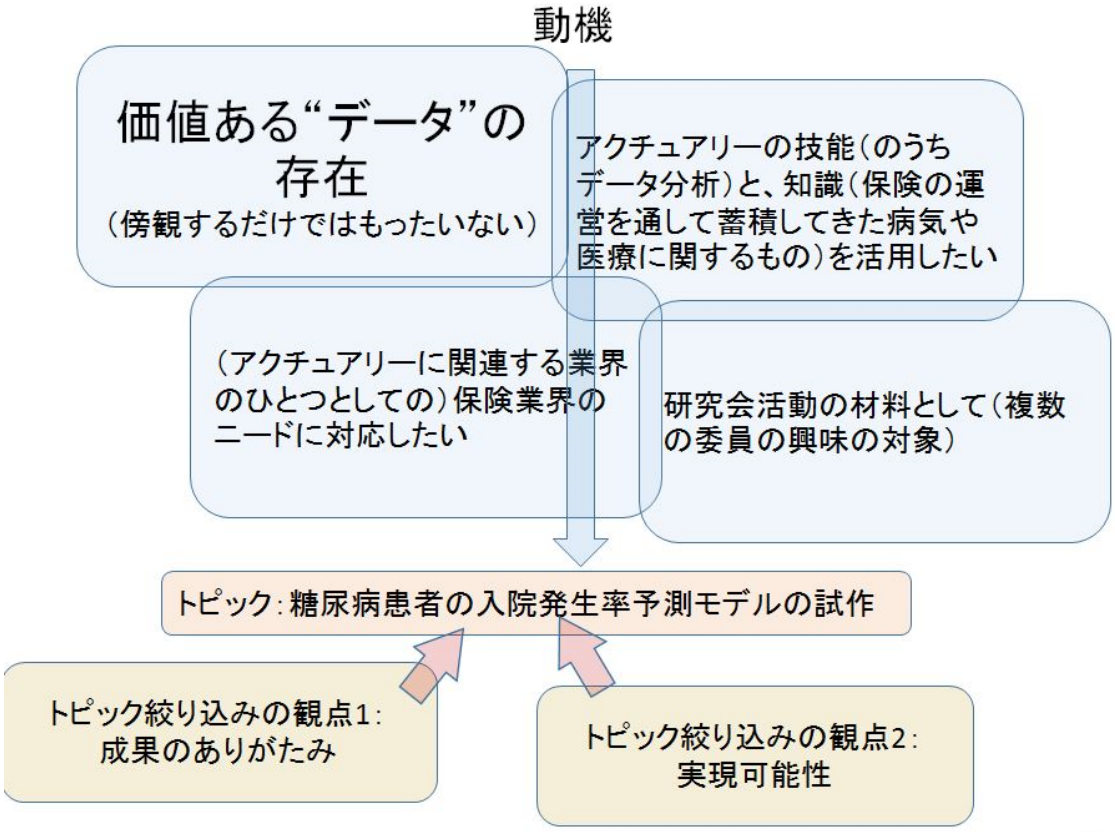
レセプト情報・特定健診等情報データベース（NDB）は、医療費適正化計画の作成、実施及び評価のための調査や分析などに用いるデータベースとして、国が、レセプト情報及び特定健診・特定保健指導情報を格納・構築しているものです。平成23年度より、医療費適正化計画策定に資する目的以外でのNDBデータの利用が認められ、有識者会議で承諾を受けた研究に対してNDBデータの第三者提供が開始されました。加えて、多くの人々がNDBデータに基づいた保健医療に関する知見に接することが出来るよう、平成28年度からは、NDBデータを用いた基礎的な集計表が一般向けに公表されています（NDBオープンデータ）。

健康・医療研究会においては、NDBは価値ある知識を掘り出すことのできる宝の山であると考え、継続的に調べています。まずは、NDBデータの第三者提供によって今までどのような研究が行われたかについて概観しつつ、NDBのデータからどのような新たな知識が得られるかを考えてみます。また、公開されているNDBオープンデータの方については、すでに集計されている情報としての限界についても考えて見ます。なお、当研究会では、大学と連携をとり、NDBデータの第三者提供による情報を基にした研究を行うべく準備を進めていますので、この研究についての考え方や進捗についても報告させていただくとともに、研究の今後の方向性についても触れたいと思います。

2

これまで健康・医療研究会においては、散発的なトピックについて、個々の委員が自発的に勉強してそれを持ち寄り、お互いに知識を高め合うということを中心にやってまいりました。けれども、去年から今年にかけては、多少まとまったトピックに取り組もうという、委員からの発案もああり、今回このようなプロジェクトにチャレンジすることになりました。

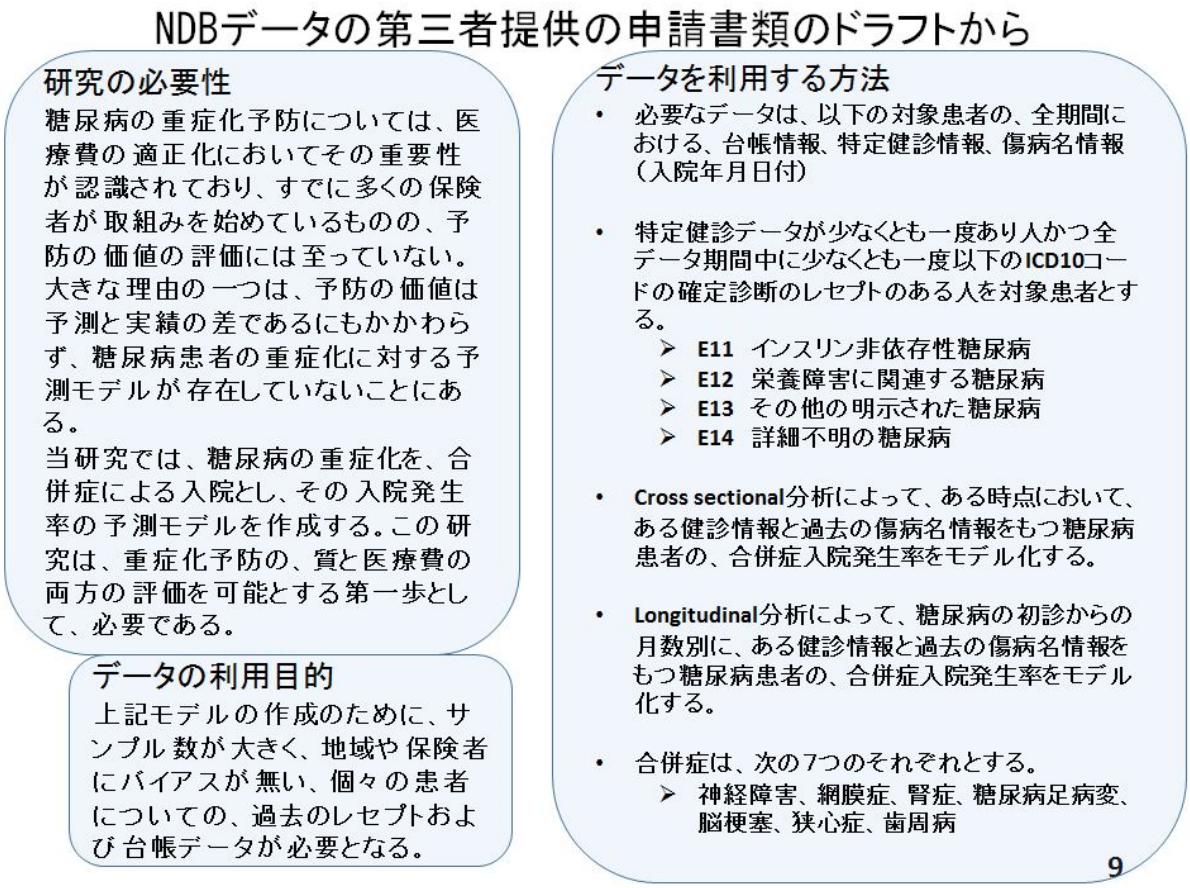
今回、ＮＤＢの利用ということに着目した理由ですが、そこに価値のあるデータがある、ということが一番の動機です。加えて、アクチュアリーの技能のうち、データ分析に関するものを実際のデータに対して活用してみたいということもありました。




それから、右上のほうの「知識」というのは、保険業界ではここのところ、レセプトデータを活用した分析をもとにした商品開発などが取り組まれてきており、そのような知識が集積されつつありますが、それらを活用したいということも動機の一つです。

具体的には、糖尿病患者の入院発生率の予測モデルを試作しよう、という課題を設定いたしました。「予測モデル」といっても概念的ですのでもう少し具体的に言いますと、すでに糖尿病の患者である人を対象に、対象者のいろいろな属性や状態、つまり年齢、性別、測定値、検査値などを入力値として、糖尿病による入院　－　これは糖尿病に特徴的な疾病や症状に対処するための入院事由（いわゆる合併症による入院）として典型的なものを選択してそれにより定義することとします　－　が発生する蓋然性を出力値とする関数が得られるようなモデルを作りたいということです。

では、このようなモデル開発には何の意味があるかということですが、ドラフトなのでまだかなり荒っぽいのですが、スライド左側の「研究の必要性」というところに書きました。




つまり、疾病は放置しおくと、場合によっては、将来に悪化し、入院、手術あるいは合併症の治療をするに至ることもあり、多額の医療費用が必要になる可能性がありますが、糖尿病は、正しくコントロールしていけば、悪化の抑制効果があるという特徴がありますので、悪化予防のためのコントロールが重要な課題です。しかし、どのような具体的なコントロールまたはアクションが、どの程度の抑制効果があるかということは、簡単にはわからないのが現状だと思います。抑制効果を評価する視点としては、たとえば、ある特定の属性・状態の人について、ある特定のコントロールまたはアクションを行った場合と行わなかった場合とで、入院になる時期や日数、回数等の期待値をそれぞれ得ることができれば、その差を、コントロールまたはアクションによる予防効果の期待値とすることができると思います。ここにおいて、特定のアクションによる予防の影響について直接測定する指標については、例えばある検査値とすれば、将来の入院等の帰結を待たずに即時性ともって把握することができるため、それらの検査値を入力すれば入院等の期待値を得ることが可能になるような予測モデルがあれば、長期の観察を待たずに期待値が得られるため、過去のデータを分析することによってそのような予測モデルを作るということにいたしました。

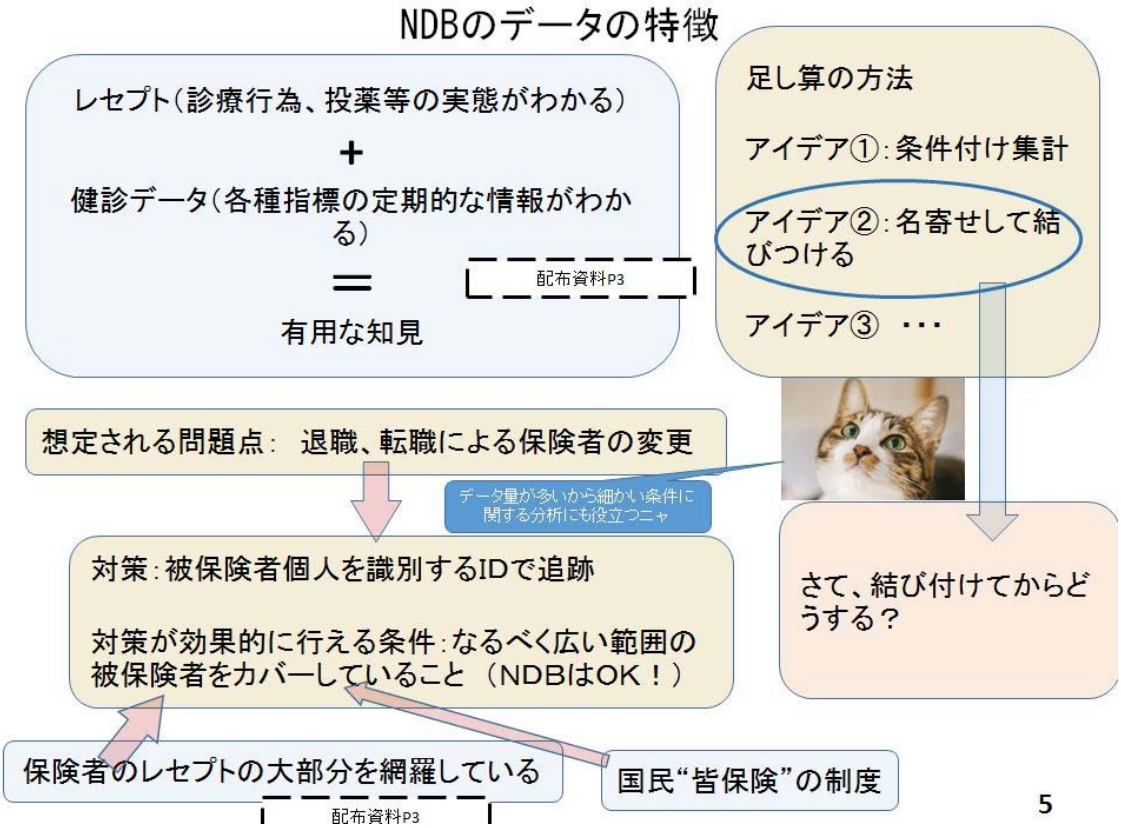
なお、このような、何かのアクションに対する効果の評価や測定は、こういった予測モデルを作るということが唯一の方法ではなく、たとえば、そのアクションを行った人の群団と行わなかった人の群団に対する最終結果をマクロに集計して、それを比較する方法が最も直接的で堅実な方法であり、このような分析は伝統的に行われていると思います。ただしこの方法だと、効果を測定すべき特定のアクションの有無以外について、同質な群団をそれぞれ一定の規模で長期に設定することが要求されるため、コストがかかります。複数の比較軸を同時にテストする工夫もあるとは思いますがいずれにしても、その他の点で同質性が要求されるか、仮にリスク調整を行うにしてもそのリスク調整にはその調整に用いる予測モデルが必要になるといった点も考慮しなければなりません。今回の当研究会の目標としている「予測モデルを作る」

ということは、そのような点にも活用できるという意味もあります。

先ほどの動機のページの左下に書いた「成果のありがたみ」ですが、このように、予測モデルができると、予防効果の指標の設定に役に立つということがあります。それに加えて、「入院発生率」を予測するということに関しては、保険業界との関連では、保健商品のプライシングや付帯サービスの開発などに役立つこともあるだろうということで、アクチュアリー会会員の方々への情報提供もそこに含まれるという意味合いもこの研究に込めております。

動機のページの右下に書いたもう一つ「絞り込みの観点」は「実現可能性」ですが、ＮＤＢデータの第三者提供を受けるというのは非常にハードルが高いようだということに関連しています。データ提供を受けることの実現可能性を少しでも上げるために、具体的に絞り込んだ研究内容にしよう、と考えました。幅広い疾病について調べたり、入院や手術や通院などの多くの事象について調べたりということを避け、糖尿病患者だけに絞って、入院だけに絞って、そしてこれが重要ですが、レセプトデータから得らる情報で分析できるものに絞り込んでいます。

ＮＤＢデータについてざっと説明いたしますと、レセプトデータは、医療機関が保険者に対して請求するための情報ですが、診療行為、たとえば、どのような診療、治療、投薬、何の投薬をした、ということが含まれています。レセプトデータの他には、特定健診・特定保健指導のデータが入っております。これらを組み合わせて分析すれば有用な知見が得られると思います。なお、これらの複数のデータセットを結びつける方法は、幾つかあると思います。




「ＮＤＢデータの特徴」のページの右側のアイデア①の、「条件付け集計」は、伝統的な相関関係を把握するための方法だと思います。アイデア②の、「名寄せして結びつける」は、個別のデータを名寄せをして、個人ごとに、過去の健診データとその後の入院の治療を有無を結びつけてから分析することができます。名寄せといっても名前でというのではなくて、ハッシュ関数を使った識別コードのようなものに変換され

たもので結びつけるのですが、このような、名寄せが分析のためにパワーを発揮します。この名寄せを行う場合に、ＮＤＢの特徴が活きます。例えば、特定のいくつかの保険組合でのデータに限定したレセプトデータセットに想定される問題点といえば、退職や転職などで保険者が変わってしまうと、そこでデータが打ち切りになって追跡できなくなりますが、ＮＤＢですと、この保険者から他の保険者や国民健康保険に移った場合にも追いかけられるということで、時間を隔てたデータ間の結びつきに関する情報があまり失われないという特徴があります。もう一つＮＤＢの利点は、ほぼ国民を網羅しているという点で、こちらの点は、後で紹介いたしますけれども、ＮＤＢを使ったいろいろな研究者の研究例に、その特質を生かした研究が見られます。このような背景がありますので、ＮＤＢデータを使うと効果的に研究ができるという点がここにあります。

ここまでで、何かご質問はありますでしょうか。

質問者Ａ　お話の中に「予防」の効果という観点がありまいたが、具体的な予防に関する情報といったものはＮＤＢの中に入っているのですか。

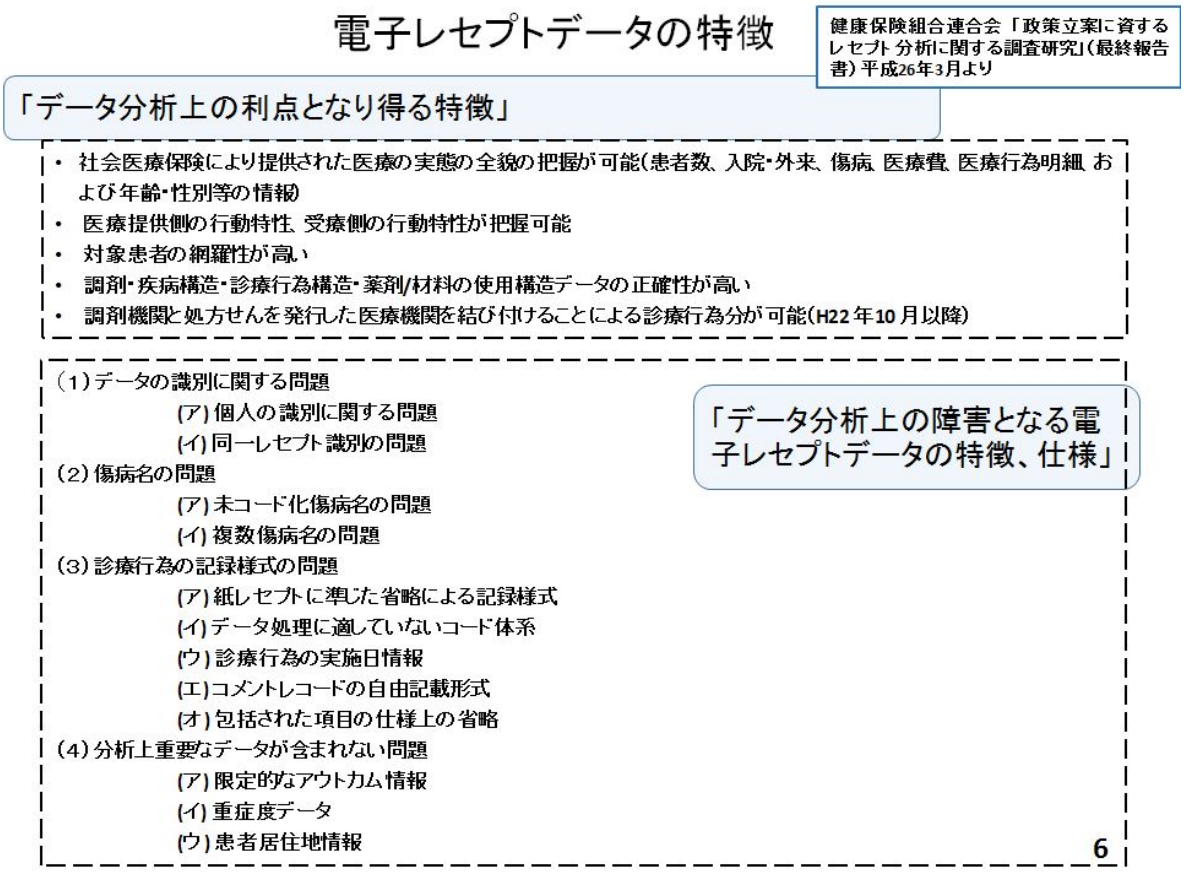
櫛引　ＮＤＢデータの中には特定保健指導のデータがありますので、ある種の予防の効果、すなわち健康指導の効果に関して、過去に、ＮＤＢデータの第三者提供によって研究された例があります。たとえば「特定保健指導の効果を指導内容別に分析」ということで、食事指導と運動指導を指導内容別に分析した研究例です。ただし、今回、当研究会で行おうとしている研究は、予防の効果を研究するのではなくて、（糖尿病の）特定の人、それは特定の属性および状態を持っていますが、その人の自然体での将来の入院率を予測するモデルを作ることにあり、そのモデルによる予測が、ある介入の予防効果を把握する上でのベンチマークに使えたらいいな、というものです。大雑把に言えば、特定の介入に予防効果があるとすれば、ベンチマークより低い入院率を示すであろうと思いますが、その差を観察によって把握できればよいのですが、長期に観察するにはコストがかかる上に、時間が経過するに伴う他の要因の混入をランダム性で中和するには一定以上の規模が必要となります。そこで、ある状態、検査値等が将来の入院率に影響するならば、そのような関係の予測モデルがあれば、入院等の帰結が判明するまでの長期観察をせずに大まかに介入のコスト抑制効果の把握に役立ちます。、あるいはそこまで言わなくても、比較対象群団を設定して長期に観察する場合でもその群団の介入群団に対する同質性を、そこに属する人の属性や状態によってリスク調整を行って同質性の精度を上げる、または、リスク調整ができるので比較対象群団の設定にことさら神経質にならなくてもよくなる、そのために予測モデルが使えるようになるならばよいな、ということです。

他に何かご質問がなければ話を先に進めますが、その前に、私のほうから質問をいたします。皆さまの中に、レセプトデータ、これはＮＤＢでなくてもよいですが、実際にデータを使って、仕事に活用された、あるいは個人的な研究に活用されたことのある方はいらっしゃいますか。若干いますね。ＮＤＢデータの第三者提供を受けてやったという方はいらっしゃいますか。この会場にはいないようですね。ありがとうございました。レセプトデータを使われたい方が会場に若干いらっしゃいましたが、最近は、保険会社で、レセプトデータやＤＰＣのデータを活用されているところもある、という話を聞きました。この中にも経験者はいらっしゃるようですね。

レセプトデータは、やはり扱いづらいデータで、話によると、結構な難物だという話は聞いております。ご存じの方も多いと思いますが、民間の業者で、複数の健康保険組合からのレセプトデータを集積して、クリーニングしたデータセットの提供や、分析あるいは分析ノウハウの提供を事業の柱の１つにしている

ところもありますけれども、そのような業者は、専門的な知識と経験の蓄積があり、レセプトデータを使いやすいように補正をしているとのことです。補正というのは、生のレセプトデータですとそのままではさまざまな欠点があるのです。

ここに「電子レセプトデータの特徴」として表示しましたけれども、データ分析上の障害となりうるデータ特徴・仕様というものがあって、これらの問題点をうまく扱わなければならない。




民間業者のデータを使用する場合には、データの補正によりある程度は使いやすいデータになっていると思うのですけれども、ＮＤＢデータの第三者提供による場合には、まずこれを自分でやらなければいけないということで、かなり困難な作業が控えていることを覚悟せざるをえません。

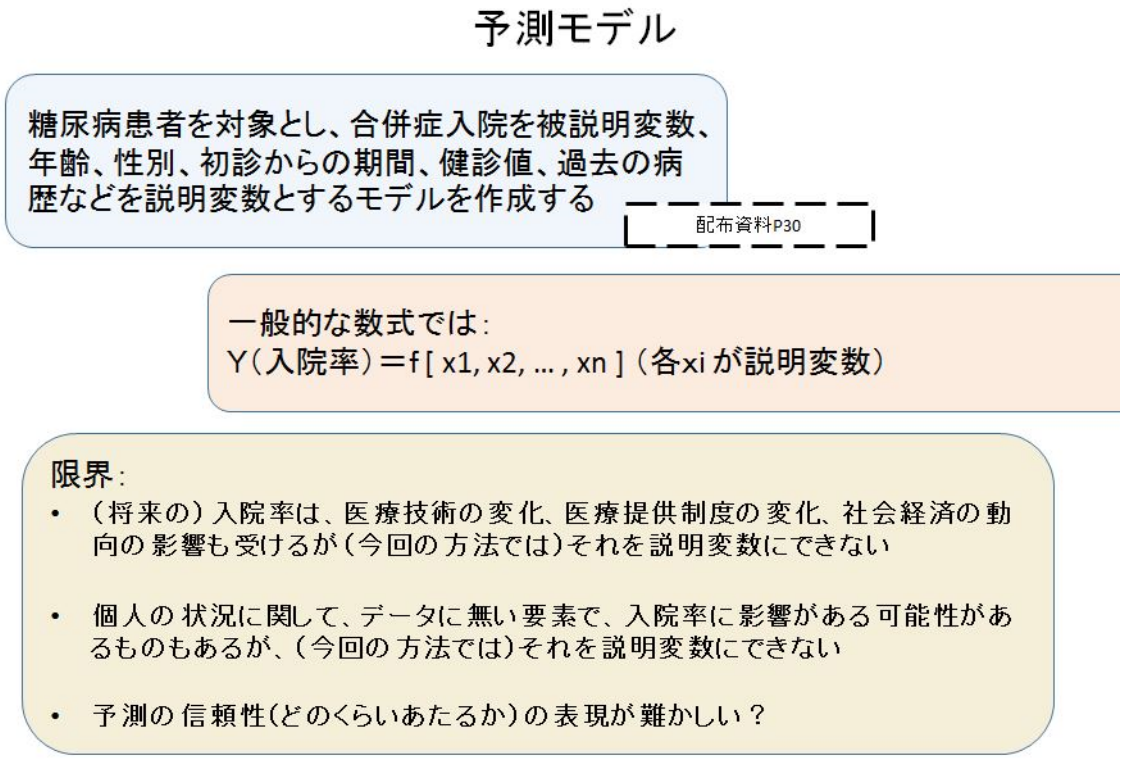
例えば、（１）（ア）の「個人の識別に関する問題」というのがあり、レセプトに書かれる名前が時によって全角だったり半角だったりする場合、あるいは旧字体と新字体の漢字の両方を使う場合だと、違う識別コードが付されてデータ上は違う人になってしまうとか、逆に、同姓同名・同生年月日の人をどう扱うかなどの問題があり、名寄せに関しても困難な課題があります。なお、厚生労働省のほうでもこのような点を課題視しており、何種類ものコードを作っておく等のの工夫を行っており、利用者がそれらを上手に使い分けることによって、名寄せをある程度効果的に行えるように取り組んできているらしいので、問題点は徐々に改善されてきているということです。

他にも、生の電子レセプトデータには重要な問題点があります。（３）（ア）の「紙レセプトに準じた省略による記録様式」というのがあり、レセプトには、一つの診療行為でもレセプト上では幾つか細かい項目に分かれるのですが、そこに点数と回数を記載するときに、本当は各項目に書かなければいけない、たとえば、３点、３点、５点で合計 11 点というように書かなければいけないのに、３点、３点、５点というものを書かずに、11 点のところにもう集約して書いて、あと全部空欄になっているようなもの、要するに

空欄になっているところは点数０として見えてしまうというような、そのようなデータがたくさんあるわけです。それは、紙レセプトで今までそのような習慣でやってきたものを、電子データでの請求に変更した場合にも、そのような習慣を変えていないということが理由に挙げられます。これに対しては、どうしようもないかというとそうでもなく、そのような空欄があっても合計欄のデータをもとに、空欄のデータを補完するわけです。何億件もあるデータをどのようにして元に戻すということは、おそらく厚生労働省のでは、自動的に補完するプログラムを開発して実行しているのだと思うのですけれども、その種の対処をすることによって、重要な問題点だったこの部分についても、かなり改善されているとのことです。

なお、このような修正をしたデータも提供が可能である、ということであり、修正前のデータに当たりたいという人は、そのようなデータを使って自分で補正することになります。

当研究会の研究の話に戻ります。予測モデルです。



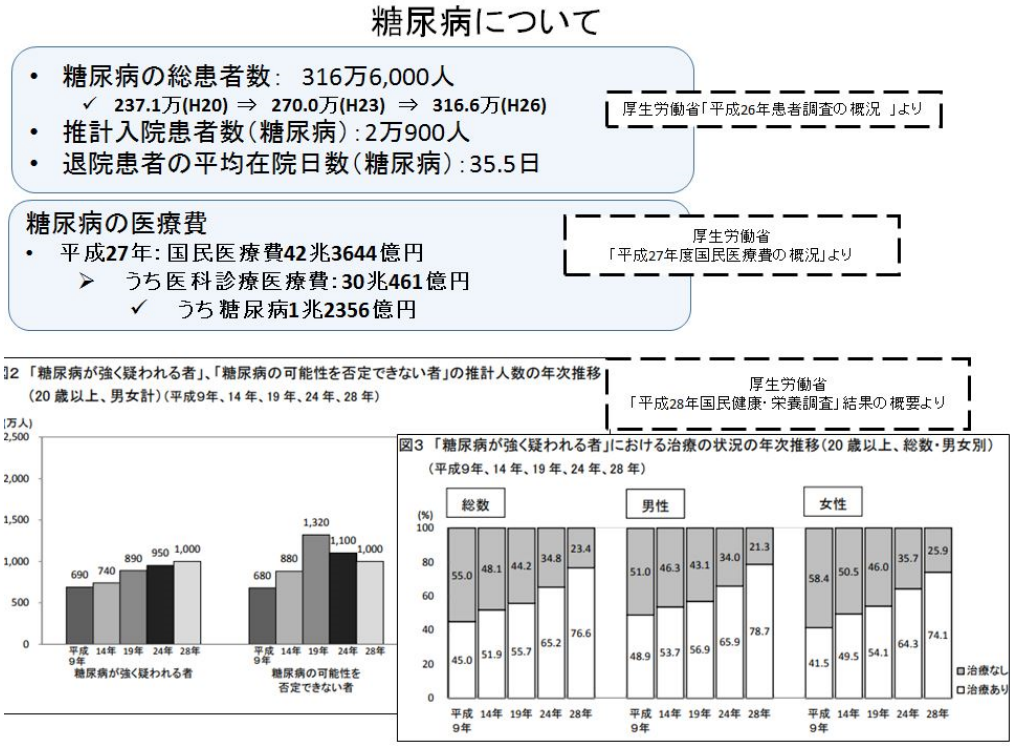

ごく一般的な概念的な算式で申し訳ありませんが、入院率は幾つかの変数の関数になっていると言うだけの話です。これは関数といっても線形とは限らないので、データからこれをどのようにして求めるのかということが問題ですが、とにかくデータ分析のノウハウを使ってモデルを作りたいと思います。適合度や情報量などの各種の指標を考慮して、使いやすい予測モデルを作るということが今回の目標です。

ただし、あくまでデータソースがレセプトデータと健診データなので、「限界」欄に記載したように、医療技術の変化や社会の変化のようなものは説明変数にできないので、全体のレベルについての将来の予測に関しては限界はあります。また、データにないものは反映できません。例えば、その家族の収入のレベル等。地域はデータにありますが、今回地域を変数に考慮するかどうかまだ決めていません。なお、地域といっても、住所に関する細かい属性、例えば、ある市町村の中の病院の近くにいる人と遠くにいる人の違いなどについては、そこまでやれるかというと、非常に困難であるかまたは分析しきれないだろうとは思います。

予測の信頼性については、その表現が難しい「？」と書きましたが、例えば、モデルの実際のデータに

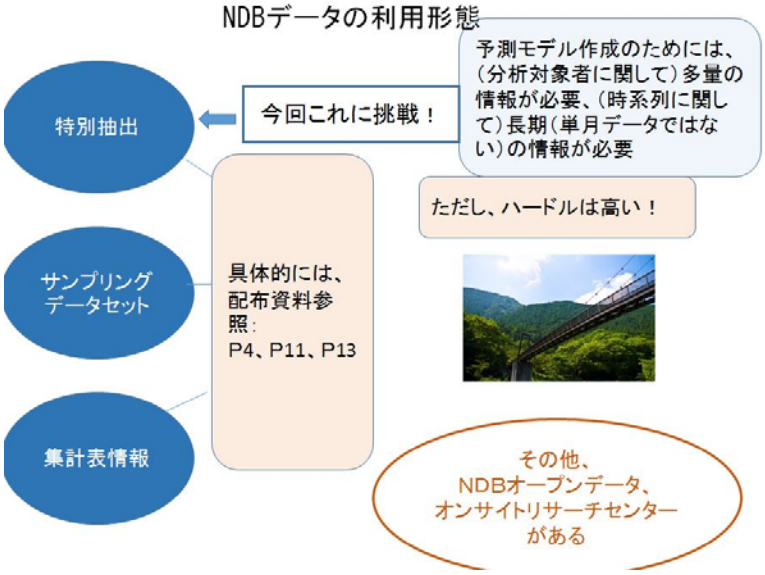
対してバックテストをして、適合度は調べることは可能だと思うのですけれども、では、将来、当たる可能性がどれくらいあるかという信頼度のようなものの表現は、この予測モデルでは難しいと思われます。統計的に集計してやる場合の形式的な信頼区間の設定とは事情が違うと思いますので「？」を付けたところです。

「糖尿病について」は、日本の状況ですけれども、かなりの患者数と、あと、予備軍として強く疑われる者が 1,000 万人ぐらいいるということで、たいへん重要な疾病と考えられます。




それから、治療をしている人がだんだん増えてきている。やはり、糖尿病をコントロールして重症化予防するというのは重要だとかねがね言われていますけれども、そのことも背景にあるのか、だんだん上がっています。先ほど「研究の必要性」の説明のときにお店したスライドでは、「予防の価値というのは予測と実績の差でもあるにもかかわらず、糖尿病患者の重症化に対する予測モデルが存在していないのが研究を阻んでいる一つ」と書きました。なお、言葉が強くて言い過ぎかもしれないのでここは修正するかもしれないのですが、要するに、この研究には、このような社会的な重要な問題に対して貢献するという意義があります、というような申請内容にしようと考えています。

ＮＤＢを第三者が利用するには幾つかの方法があって、今回は「特別抽出」という、研究者側からの具体的なデータ抽出リクエストに基づいて提供を受ける、ということに挑戦しようと思います。一番難易度が高いものです。




その他のものでは、比較的難易度の低いものにはサンプリングデータセットといって、あらかじめＮＤＢのほうでサンプリングベースのデータセットが作ってあるというものです。ですので、それを使いたいという申請をすれば、そのデータセットが貸与されることができるということで、申請に対するいくつかの要件が多少緩和されているということです。また、「集計表情報」というものもあって、それを使うというのも一つの方法です。集計表情報に関しては、その後、ＮＤＢオープンデータというものができているのでそれを使うという方法もありますけれども、後ほどそれについては説明いたします。

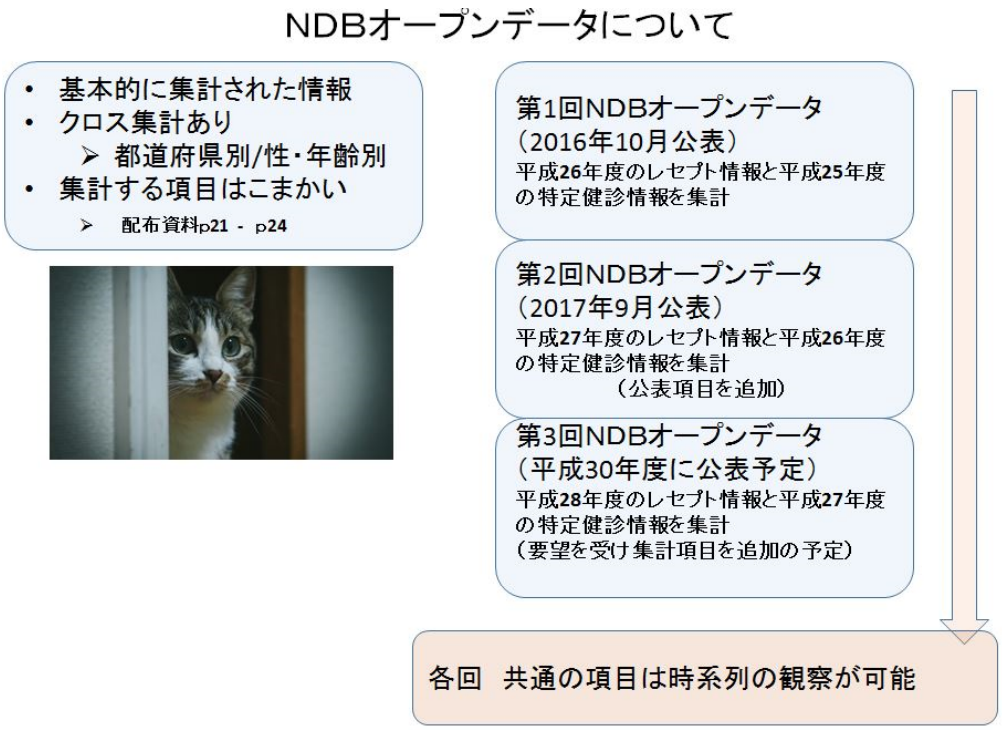
今回検討している、「特別抽出」はハードルが非常に高いということは認識しております。ＮＤＢの提供を受けようと思ったけれどもうまくいかなかったなどの話を時々耳にしますし、「実際のところ申請を通すのがすごく大変なんだよ」というは話は聞いております。したがって今回の取り組みにおいて、最初はなかなかうまくいかないかもしれません。要件が曖昧すぎる、抽出条件が緩すぎてデータ量が多すぎる、あるいは、研究内容とのバランスが悪いなど、いろいろな指摘を受ける可能性があるかもしれないですが、そのようなものを１つ１つ解決していって、まずは申請を通して第三者提供を受けてたいと思っています。そのような研究をいちど実現できれば、その経験を次の研究の企画に活かすことができると考えております。そして将来的には、私の権限を越えますが、日本アクチュアリー会自体がイニシアチブを取ったタクスフォースで、ＮＤＢを利用して研究できるよう、そのような環境づくりに少しでも役に立ちたいと思っています。

今回は、健康・医療研究会でイニシアチブを取って、この第三者提供の申請に挑戦しようとしているのですけれども、正直なところ、健康・医療研究会の代表者の私の名前で申請をしても、承認される可能性は小さいと考えています。そこで、研究の手続きに関して工夫をしております。今回は、大学の研究室の先生と協力し、その先生の研究活動として申請をしてもらおうと思っております。研究の実態についても申請内容に一致すべきなので、ＮＤＢデータを利用した研究を行ってその研究を公表する主体は当該大学の当該研究室の当該先生であるということになります。当研究会は、研究計画作成と申請の段階では、その研究目標や方法論の企画などについて先生と連絡を取りながら方向づけに協力していくというスタンスです。提供を受けての分析の段階では、研究会の委員の 1 人が当該大学の受託研究員になって、当該研究室の内部で研究を行います。研究結果は可能な限り具体的に、集計データ、各種指標、分析の手法や予測モデル等、公表される方向で行きたいと思います。その後、再び当研究会にて公表された結果について、

その有効性、応用の有用性、当該データ分析による予測モデルの作成の理論についての議論等、当研究会でコメントするような二次的な研究を行って大学における最初の研究と接続しようというのが今回の企画です。

また、将来、先ほどアクチュアリー会としての研究タスクフォースの可能性うんぬんという話をしましたが、そういった機会に、今回お願いしている先生にタスクフォースに入っていただくなど、そのようなことで発展させるということも考えられます。なお、アクチュアリー会を主体として実施するためには、アクチュアリー会に、鍵の掛かる、セキュリティが万全なコンピューターを用意してあるかどうか等の要件を満たすことも必要となりますので、実現へ向けての課題はあるでしょうが、将来への布石になればよいと思います。

先ほどそのオープンデータについて話すを予告しましたが、オープンデータとは、今でも誰でもアクセスできるデータセットです。Ｗｅｂサイトで公開されています。




これは先ほどお話ししました特別抽出とは違って、一件一件の個別データではなくて、すでに集計された情報なので、今回の私どもが考えている予測モデルを作るには不十分な情報であるため、今回の研究はこれでは行いません。オープンデータには、都道府県と性・年齢別というクロス集計があり、各変数についての細目は細かく設定されており、それはそれでかなり便利なデータで、これがオープンになったときにかなり話題になったのですが、他の変数との相関を分析することはできません。

来年の春ぐらいに第３回のＮＤＢオープンデータが、さらに内容も充実して公表される予定です。３回目になるので、時系列に関する分析もある程度可能になると思います。皆さんもご興味があればこの情報を活用して創意工夫した分析をしてみてはいかがでしょうか。

## NDB今後の動向

オンサイトリサーチセンター：稼動開始のための準備

オンサイトリサーチセンターの現状について

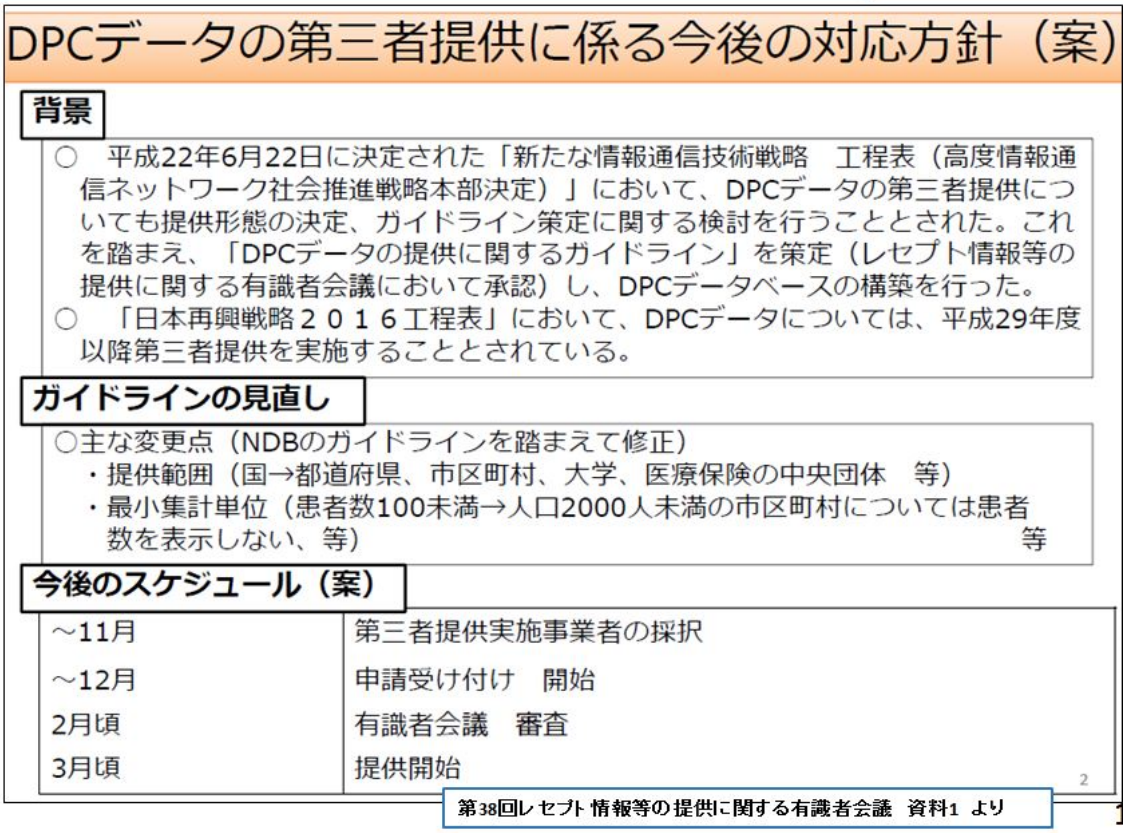
配布資料 p38 参照

DPCデータの第三者提供の方針

今後の動向では、オンサイトリサーチセンターという話もありまして、現在東京大学および京都大学での試行が実施されており、本格的な運用開始に向けて準備中のものですが、これは、厚労省が大学の設備を利用してＮＤＢの情報を利用できる環境を整備し、研究者はそこに行ってデータの利用をして研究し、許可された出力だけを持って帰るというような形で行われるものです。この場合においても利用については審査があるのですが、当然のことながら、レセプト情報等の利用場所や保管場所及び管理方法に関しては、審査の対象にはしないことが検討されています。アクチュアリー会としての研究タスクフォースの可能性との関連で考えて見ますと、この点においてハードルも低くなると考えられます。

それから、第三者提供が検討されている他の種類のデータについての話に触れておこうかと思います。ＤＰＣデータです。

DPCデータの第三者提供の方針



DPCデータの第三者提供に係る今後の対応方針（案）

**背景**

○　平成22年6月22日に決定された「新たな情報通信技術戦略　工程表（高度情報通信ネットワーク社会推進戦略本部決定）」において、DPCデータの第三者提供についても提供形態の決定、ガイドライン策定に関する検討を行うこととされた。これを踏まえ、「DPCデータの提供に関するガイドライン」を策定（レセプト情報等の提供に関する有識者会議において承認）し、DPCデータベースの構築を行った。

○　「日本再興戦略２０１６工程表」において、DPCデータについては、平成29年度以降第三者提供を実施することとされている。

**ガイドラインの見直し**

○主な変更点（NDBのガイドラインを踏まえて修正）
- 提供範囲（国→都道府県、市区町村、大学、医療保険の中央団体　等）
- 最小集計単位（患者数100未満→人口2000人未満の市区町村については患者数を表示しない、等）

等

**今後のスケジュール（案）**

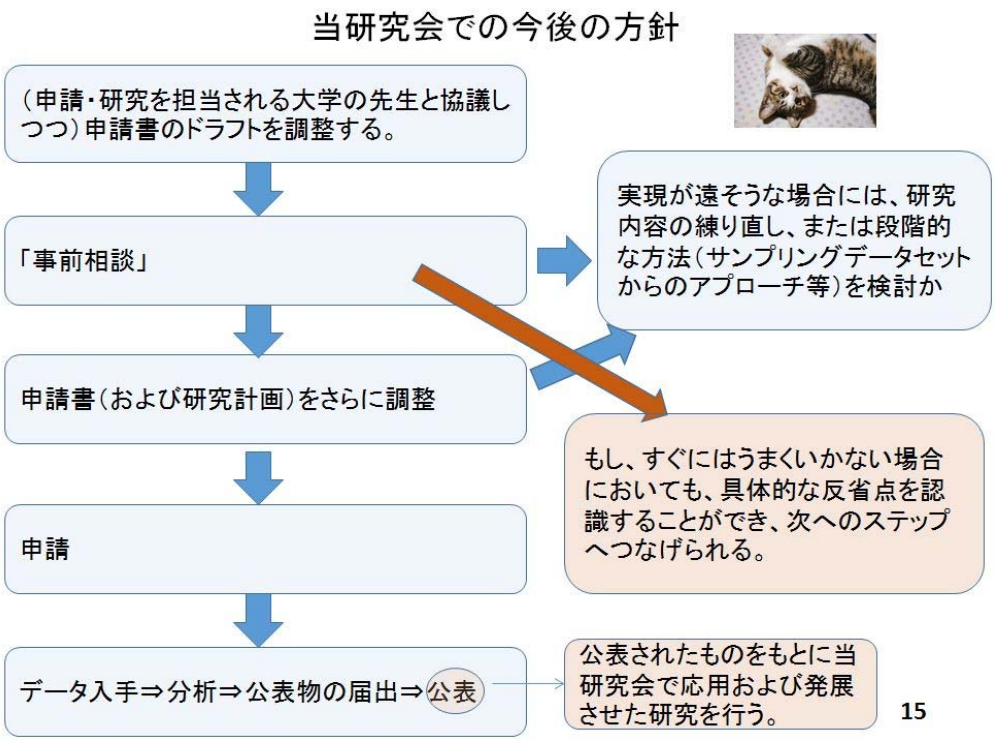
| | |
|---|---|
| ～11月 | 第三者提供実施事業者の採択 |
| ～12月 | 申請受け付け　開始 |
| 2月頃 | 有識者会議　審査 |
| 3月頃 | 提供開始 |

2

第38回レセプト情報等の提供に関する有識者会議　資料1 より

ＮＤＢデータのレセプトデータの中にもＤＰＣレセプトも当然含まれていますが、ここでいうＤＰＣデータとは、ＤＰＣレセプトで保険者に請求する病院が、レセプトとは別に厚生労働省に提出している患者の臨床情報および診療行為の詳しい情報が含まれている電子データです。このＤＰＣデータについても、第三者提供を検討しよう動きがあります。ただし、個人や医療機関を直接指し示す項目意外の項目に、それらを特定または推定できる情報が含まれているため管理が難しく、現段階では集計済データの第三者提供の検討になっております。しかし、レセプトデータにはない情報が含まれているデータですので、これは将来に期待したいと思います。本日はＤＰＣデータについては深入りした話はできません。

今回の研究での今後の進め方ですが、まず第三者提供の申請のプロセスとしては、事前相談というものをいたします。

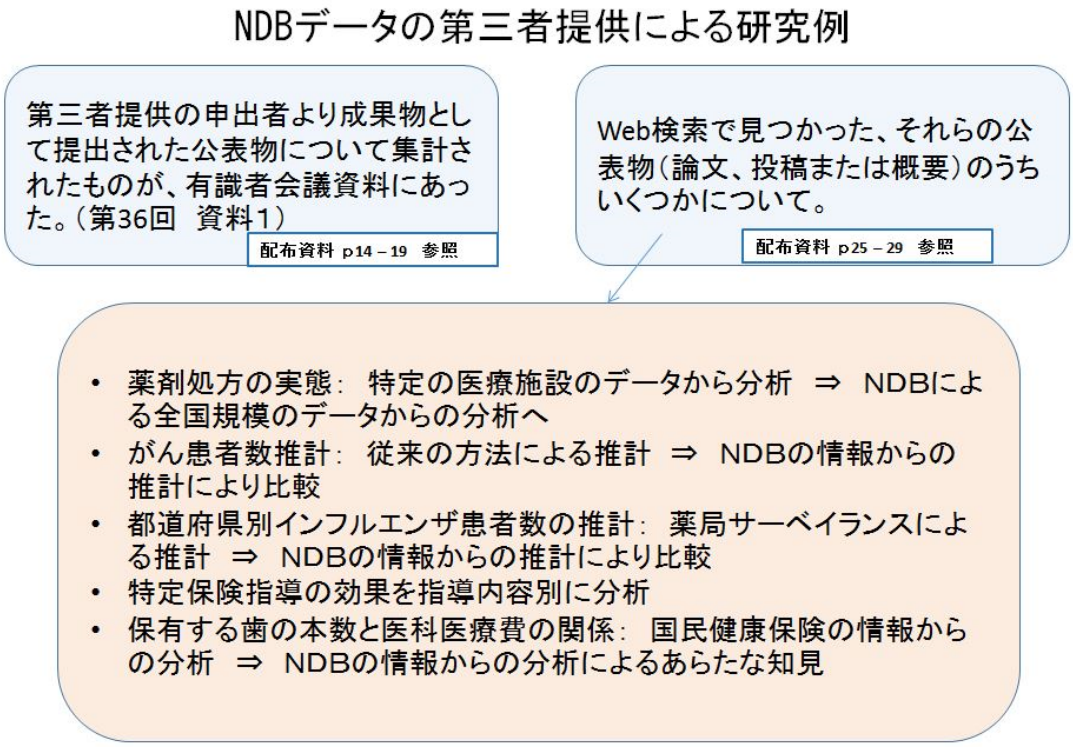



事前相談のときに形式的なチェックをして、その後、申請書の提出、スライドには書かなかったですが、その申請については、有識者会議で審査されて、オーケーかだめかという結果が出されます。恐らく、事前相談においては、種々の注文が付くだろうど思います。データの抽出の仕方が曖昧である、あるいは、データ管理のセキュリティ態勢が本当に万全なのかなど。もしかしたら、そのような問題点を解決してからもう１回相談に来て下さい、というような話になるかもしれないのですが、それはそれで解決すべき問題点が何であるかということが具体的にわかってきますので、そのステップによって、ＮＤＢのデータを使って研究するためのノウハウを蓄積することにも意義はあると思います。

少し話を変えます。先ほどご質問への返答で一例をお話をしましたが、過去、ＮＤＢデータの第三者提供によってどのような研究がされたかということについて、コメントしたいと思います。まずは全体像から。資料集を１４ページから１９ページを見ていただきますと、この時点で 105 件出ています。提供依頼申出者の名前を見てみると、同じ人が複数回出ています。１回経験すると２回目以降は、その経験が活かせることが想像できますね。この一覧は、１年前ぐらい前の時点の資料から、もとのデータとしては１年半ぐらい前かもしれないですけれども、その結果が、公表されたものだけが記載されていますので、現時点での第三者提供を受けた累計件数は、105 件より多くなっていると思います。

なお、研究結果を公表する際においても、公表する内容について、事前に厚労省に提出をしてその内容で公表してよいかについて確認を取らないといけない。場合によっては、その公表内容について、有識者会議で公表の可否についての審査が行われるというプロセスになっておりますので、研究結果がまとまったからといってすぐ公表できるわけではく、公表まで時間がかかるということもありうるということです。

具体例としては、５つをここで取り上げてみました。スライドにはごく簡単に間と待てありますが、資料集のほうには、２５ページから２９ページにかけて、それぞれの公表資料の研究要旨の部分等からもう少し具体的に内容をピックアップして記載してありますので、別途ごらんになってください。




１つ目は、薬剤処方の実態に関する研究です。どのような薬を、どの程度、どのタイミングで処方したかというようなことは、製薬業界やお医者さん、病院にとっても重要な情報ですけれども、今までは、特定の病院を幾つかピックアップした調査を行い、それで分析をしていたということです。ただし、サンプル的な病院から取ったのでは、バイアスもあるし、データも少ないので、誤差も大きいかもしれないという心配がありました。それに対してこの研究ではＮＤＢによって全国規模のデータから分析できるようになったことから、付加価値が得られています。また、これは、ＮＤＢ第三者利用のわりと初期のころの研究なので、ＮＤＢを利用するときの苦労といいますか、問題点や特徴についての分析もしてあり、参考になります。

２つめは、がん患者推計です。このあたりは保険会社の方にとっては興味のあるところだと思います。がん患者数を推計する方法については、患者調査からとってくるのがもっともポピュラーな方法ですが、これは特定の月での特定の施設で調査したものから推計されたデータという特徴があります。従来からの方法としてはもう一つ、有病率を推計するという方法で、がん患者の生存率と死亡率、死亡数から計算して推計するという方法があります。さらには、患者調査の推計と有病率の推計を比較して、相違のある二つの数字を見ながら、推計方法の特徴を踏まえつつ、調整して使用したりするような方法もあるようでしたが、この研究は、ＮＤＢから、レセプトデータを使って推計してみよう、という内容です。結論では、レセプトデータからの推計と、従来の方法による推計との間には、結構な差がある、特に高齢者の推計において、というようなコメントが書かれています。

３つ目は、都道府県別インフルエンザ患者数の推計の精度を上げるためにＮＤＢを利用する例ですが、ここまでの３つは、マクロベースでの状況の把握について、全体を網羅しているというＮＤＢの特徴を生かした分析になっています。

のこりの２つは、ＮＤＢの特徴としてデータ量が多いという点を特に意識して活用しています。データ量が多いということは、より細かい分析をする場合においても、その材料が十分に得られるというような利点につながります。４つ目の特定保健指導の効果については、それまではの研究では指導の内容別行わ

れたものはあまりなかった、論文には書かれています。それを、この研究ではＮＤＢのデータを活用して食事指導と運動指導の効果を分けて検討したということです。

５番目の例では、歯の本数と医療費の関係です。保有する歯が多い人ほど、医科医療費が少ないという個別の研究は従来からあったようですが、いずれも、保険者を限定した範囲の情報からでの研究だったところ、今回、ＮＤＢによる大規模データで研究し、そのような結果が再確認できた、という内容です。

研究者の何人かが書いているコメントの中に、ＮＤＢの有用な活用方法としては、まずは、今までの個別の研究、サンプリング的なデータで研究されてきたものを、再確認したり、その結果を検証するような研究に使うような方法が、取り組みやすいとのことです。その理由としては、この第三者提供を受けるときに、データをどのような条件で抽出し、どのような分析を行い、分析の結果、どのように結果を表現するかということろまで含めて、申請段階において具体的に決めておかないと審査を通すのが難しいため、得られるであろう結果について、その振る舞いがある程度把握できるものがやりやすいでしょうとのことだと思います。現に、ガイドラインやマニュアルには、探索的な研究は、審査で通りませんということが書いてあります。アクチュアリーでデータ分析に長けている方は、むしろ探索的な研究のほうが好きだったり得意な方が多いと思います。データをまず手に入れて、それをあれこれいじってみて、構造が見えてくるに伴い、次はこの観点から分析しよう、というように。最初から結果を表現する構造が分かっていて数値だけがアウトプットであるようなものよりは、データの海に飛び込んで宝を発見してみるというようなことが得意だと思います。そのような設計がＮＤＢの第三者提供を受けての研究ではできないのため、利用するデータ項目の範囲、変数の体系、集計・分析方法等をあらかじめ定義しておかなければいけないというところが、難しいところだとは思います。

それから、研究内容に対して、提供を要求するデータが必要十分なものであるということが審査の要件になっていて、例えば、ある一つの項目について無条件に、無条件にとは、全部抽出しましょうというのがあるものはだめです、と注意がなされています。ですから、必要な限りにおいて、データの抽出を絞る必要があるということです。したがって、絞るためには、そのデータの内容についてある程度知っていなければいけません。ですので、ビッグデータを使って、機械学習をして、予見していなかった知識を拾っていこうというアプローチには、ＮＤＢデータの第三者提供の場合には、充分な活用の機会を得ることができないかもしれません。これは、私の現段階での印象をいうこともありますが。

質疑応答の時間に移ろうかと思いますので、コメント・質問がありましたら、お願いします。

質問者Ｂ　発表、ありがとうございます。

２つほどあります。１つは、これは、やや長めの研究になるかと想像します。2026 年の東京大会、ＩＣＡに向けて７～８年あるとは言うものの、意外にあっという間にその時は来ますので、ぜひとも、このプロジェクトがうまく進んで、できましたら、ＩＣＡ用に何らかの成果を提供できるように、そのことを視野に入れながら進められることを希望したいと思います。

もう１つは、テーマ選定に関したものです。糖尿病の入院ということなのですけれども、言葉として、「入院」や「重症」などの言葉が出てきたのですけれども、私は、実は、先ほどのグラフに平成９年から平成 28 年まで時系列がありましたけれども、患者数のところも治療の白いところも、私自身が数に入っています。約 20 年間の糖尿病患者としての知見もあり、多くの資料も調べましておりまして、それで、ガンと違って、ステージ１～４のようなレベルが糖尿病にはあまりなく、合併症が非常に多く、主に腎臓や、目の網膜や、歯の歯周病との関係や、神経系など、いろいろ関連は言われていますが、糖尿で入院となるもの

では、人工透析なども多くある、そんな気もしているのです。糖尿病には、ステージのような概念がないのではないかということ、それと、合併症が多岐にわたることを考えますと、なかなか分析がしにくいと思います。このような糖尿病の特質を考えると、ＮＤＢデータを使って何か分析するときに、なかなかうまくいくテーマ選定なのかどうかというところは、若干気になるところです。あと、主にお医者さんによるデータになりますのと、そのようなものは、アメリカにおける研究などがそこそこあるような気がしています。それで、糖尿病ということに関しても、アメリカにおける行研究のようなものが参考になると、分析のヒントになるような気も少ししております。

すみません、まとまりませんけれども、ぜひともＩＣＡの 2026 に向けて、このプロジェクトがうまくいくように願っております。

櫛引　ありがとうございます。そうですね。ＩＣＡの件については、この研究活動を活かした何らかもので貢献できればよいと思います。

それから、入院に関して、まさにこのページに、これは患者調査の概況なのですけれども、この入院の数値には、糖尿病と書いてありますが、例えば、他の病名での入院の数に入っている患者もあって、糖尿病患者の方の一部はそちらのほうに入れてあるなどということもあるので、これが糖尿病による入院の人数の実態を正確に表しているかというと、定義の問題とはいえ一概にいえない可能性があると思います。

そのようなこともあって、今回かなり割り切って考えておりまして、代表的な合併症を参照して、糖尿病を持った人が、合併症で入院したものを、今回のいわゆる「糖尿病による入院」と定義をしています。そのような定義に基づく分析を行ってみて、かつ、その定義が、実態把握に対してどの程度有用性をもった情報につながるかという点については、別途考察をしていきたいと思います。

それからもう１つ、アメリカにおける先行研究を参考にする件については、この研究の企画段階でできれば理想的ですが、少なくとも、データ分析結果が公表された後段階での、当研究会によるそれの考察を行うときに、そのような海外の研究と比較して論考を行うということも考えられます。そのような意味でも、データ分析によるモデル作りという第一段階と、当研究会でのそれに基づく検討とを結びつけるよい材料にすることができる思いますので、今おうかがいしたお話を参考にして進めていきたいと思います。

司会　まだ何かありますか。では、最後に。

質問者Ｃ　途中のご説明の中で、ナショナルデータベースの第三者提供はハードルが高い、大変だ、うまくいかないというような話が何回も出てきたのですが、それはどこが大変なのか、そこを知りたいのです。セキュリティの要件が厳しいという話や、それから、データの定義が曖昧ではないかなど、そのような点が厳しいという話は分かったのですけれども、それ以外はどのような点が大変なのでしょうか。

櫛引　今のところ、把握している大変さは、その２つのところです。もしかしたら、まだ知らない難しさがあるかもしれません。なお、第三者提供による研究を行うことができる、代表者が所属する団体については列挙されておりまして、大学はそれに入っています。公益法人については、医療の質等に関しての向上をその設立目的に含んだ公益法人であることが要件なので、アクチュアリー会自体はそのような趣旨を直接・明示的には含めていないので、今のところは難しいかもしれないですね。もちろん間接的には貢献しているとは言えるでしょうけれども。そのような意味で、所属する団体自体にも要件というものがあり

ます。それから、複数の研究施設にまたがる研究は、セキュリティの面におけるハードルが高くなるということもマニュアル、ガイドラインにありますので、研究活動自体を共同で行う共同研究というのも、高いハードルを越える必要がでてきます。

質問者C　ありがとうございました。

司会　それでは、以上をもちまして、セッションG「レセプト情報・特定健診等情報データベースを利用した研究について」を終了いたします。

櫛引　ありがとうございました。

司会　皆様、櫛引様には、いま一度大きな拍手をお願いいたします。