

特別連載

— 日本のがん対策の新しい動き —

— 科学的根拠に基づいたがん対策を進めるために —

関連データのリンクによるがん対策の
情報インフラ構築へ向けた試み

国立がん研究センターがん対策情報センターがん政策科学研究所

東 尚弘

はじめに

科学的な根拠に基づくがん対策を進めるためには、実態の把握と施策と目標達成度の継続的なモニターが必要である。しかし、そのためには良いデータ源が必要である。幸い、様々なデータが独立に様々な目的で収集されており、それらを組み合わせることができれば、有用なデータ源になり得る。実際、英国ではこの考えのもと、National Cancer Intelligence Networkが2007年から設立され、様々なデータ源をリンクして解析することが行われている¹⁾。その設立時、国のがん対策長官(National Cancer Director)であったマイク・リチャーズ氏の言葉がホームページには引用されている。「きわめて単純に、我々は2012年までに世界で最高のがん情報システムを望んでいるのだ。」

データシステムの構築を目指しているのは英国だけではない。米国においても情報システムの構築は進んでいる。がん登録の一つであるSEER(Surveillance, Epidemiology, and End Results)と、連邦政府の管理する公的保険で65歳以上の国民が主対象のMedicareとのリンクは従前より行われてきた。最近ではさらにそのようなデータ源

を広げ、医療政策や臨床、患者支援を円滑に進めるために「学習する医療システム(Learning Healthcare System)」を構築するのが必要不可欠、とする米国医学研究所(Institute of Medicine)のレポートが出版された²⁾。その推奨を受け、臨床腫瘍学会(ASCO; American Society of Clinical Oncology)ではCancerLinQというシステムを開発し、医療の質の検証や研究用データの整備が進められている³⁾。

わが国の現状はどうか。2013年12月6日に「がん登録推進法」が成立した。これで、やっと全国のがんの罹患数が正確に把握するための全国がん登録の土台ができたところである。これから先を走る欧米先進国に追いつくことはできるだろうか。先だって同年5月に法律が成立したマイナンバー制度も、医療分野に使えるかどうか不透明であるなど、個人情報保護への懸念が非常に強い我が国では、何が可能なのか、まだ模索が続いている。

1 データ間のリンクの可能性

そのような中で、研究ベースではあるが、全国のがん診療連携拠点病院で行われている院内がん登録と、診療報酬制度に合わせて収集され「DPC

(Diagnosis-Procedure Combination) 導入の影響評価にかかる調査」における E ファイル, F ファイル (あるいは EF 統合ファイル) をリンク可能な形で匿名化して収集する試みが開始された。2012 年度はがん臨床研究事業「がん対策における管理評価指標群の算定における既存データの可能性に関する研究」で 13 施設における 5 がん (胃・大腸・肺・乳・肝のがん) の症例について、2013 年度は対象を拡大し、がん研究開発費「がん臨床情報データベースの構築と、その活用を通じたがん診療提供体制の整備目標に関する研究」で 5 がん+前立腺がん、子宮頸がんについて、データを収集している。DPC の対象期間は、診断年の前年 9 月～診断年の翌年 12 月まで (例えば 2010 年症例ならば、2009 年 9 月～2011 年 12 月) のデータを抽出した。さらに、自施設治療例を対象とするため、院内がん登録の症例区分で対象症例を絞った (症例区分コード 2~4) これが成功すれば、やっと米国の SEER-Medicare のデータと同じ解析や研究が可能になったといえる。希望的な見方をすれば、日本版 Learning healthcare system への第一歩である。

1 ●リンクの目的

これらの研究は、もともと標準診療実施率を質の指標 (Quality Indicator) として、効率的に標準実施率の算定を行うことを目的に行われたものであり、都道府県がん診療連携拠点病院連絡協議会・がん登録部会 の場で募集を行ったことから、「がん登録部会 QI 研究」と呼ばれている。なぜリンクが必要かという、標準診療実施率を検討するには、院内がん登録、DPC それぞれに利点欠点があるからである。院内がん登録は、がんの種類・組織型やステージなど情報として登録されており、研修を受けたがん登録実務者が各拠点病院に配置され登録にあたっていることや、国立がん研究センターからの施設訪問などがあり、一定の質が確保されていることが利点である一方、行われた診療行為については初回治療に限られ、また手術・放射線療法など各治療の有無だけであるなど限界がある。逆に DPC は、行われたことが月ごとに全て記録されているものの、がんのス

テージなどの情報についてはその質を担保する仕組みが十分とは言えない。そこで、これらの両方のデータがリンクできれば、お互いの欠点を補い利点を活用することができると考えられた。

2 ●実際の方法

この試行において最大の特徴は、個人情報保護のため院内がん登録と DPC は匿名化される際に共通の番号で匿名化することである。当然、個人情報が施設の外に出るはいけない、というのが最低限守るべき条件との認識で行われた。院内がん登録はこれまで年に 1 度、匿名化の上国立がん研究センターに集積されてきた。同じ匿名化 ID を各施設で DPC につけ、さらに、各参加施設から、国立がん研究センターに対してこの集積されたデータから自施設分を研究班へ抽出提供を依頼し、研究班で提供を受けたデータ同士のリンクをして解析することを行った (図 1)。

各施設内での処理は全て専用のソフトを使って行った (図 2)。院内がん登録提出の際に、匿名化 ID (登録の連番) と患者 ID 番号などの個人情報との間の対応表を生成し施設内に保管している。その対応表をもとに DPC のデータを匿名化するのを基本と考えた。DPC のデータが欠落している部分 (特に外来など) には通常のレセプトで代替して同様に匿名化することとした。DPC やレセプトのデータを抽出する際には、対象となっているがんの患者以外の患者データも含まれている。これらを除いた対象患者のみのデータとするため、国立がんセンターの集積データから対象患者を部位などの条件から抽出し、それらの匿名 ID だけをリストとしたファイルを送付した。データ抽出も一括して研究班で行うことで、データ抽出の細かい手順によるばらつきを極力排除するようにした。

専用のソフトウェアは、この対応表と DPC データを読み込み、対象症例のみに対する DPC データ出力を匿名化・症例抽出後に出力する。このように施設内で行うのは、全てのデータを読み込んで処理をかけることだけである。過程を全て自動化することで、特に専門的な知識がなくても全ての匿名化処理が病院内で行えるようにした。



図1 院内がん登録の利用の手順

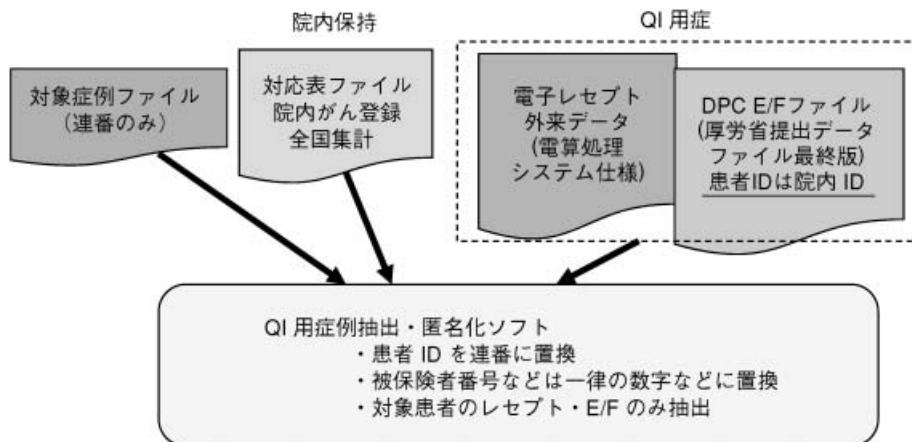


図2 匿名化ソフトの動作

さらに、このデータを送付する過程でのさらなる安全性を確保するために、匿名化したDPCデータは自動的に暗号化処理が施され、研究班で保有する復号化ソフトを通してのみ復元ができるように設計した。

2 これまでで判明した課題と解決

初年度のパイロットは特に大きく問題はなくデータの収集は行うことができた。2年目に2011年症例についてのデータ収集活動は、原稿執筆現在、100施設ほどのデータの収集が完了しているが、実は初年度ではわからなかったものも含め、いろいろな問題が判明した。そのいくつかと解決策を以下に述べる。

1 ●施設の保有するDPCファイルが既に匿名化済

院内がん登録提出時に作成される対応表は、院

内がん登録の匿名化ID(連番)と、院内の患者ID番号を対応させている。一つの大きな問題は、今回処理の対象となったDPC調査のデータは既に匿名化がなされているということであった。そのため、対応表をそのまま適用すると対応する症例がマッチしないという問題が起きる。当初はDPCファイルを匿名化の前の状態に戻してもらうことを依頼していたが、逆にDPC調査での匿名化の方法に従って対応表上にある本物の患者ID番号を匿名化した上で、処理を行う方が効率的であると考えられた。この問題があった施設に依頼するとすぐに対応してもらえて解決することが多かった。

2 ●レセプトの患者ID番号はそのまま、DPCは匿名化済

この問題は少し応用編である。レセプト上の患者IDは、通常請求目的にしか使われないために、患者IDを含む個人情報匿名化されていない

い、そのため対応表を前項のように匿名化処理をしてしまうと、レセプト上位の患者IDはマッチしなくなってしまう。ここはやはりDPC調査のファイルの匿名化を元に戻してもらうしかない、というのが実情であった。しかし、月ごとに分離されて保存されているファイルをいちいち修正するのは大変な作業であるし、そのような処理がすぐに行える施設と行えない施設がある。あるいは、レセプトとDPCで別々に匿名化処理を行ってもらう方法もある。しかし、それではDPCだけ、レセプトだけに存在する症例もあることから、データチェックで全対象症例の存在などをチェックしているため処理がそこで止まってしまう。チェックを外すことも可能だがエラーの入ったファイルが生成される可能性が高い。

そこで、ソフトを改修し、対応表を二つレセプト用とDPC用に区別して使えるようにした。レセプトによる一部期間の代替を行わない施設は、一つの対応表で良いため、その二つは選択可能にした。これで、レセプトもDPCも同時に処理ができて一括してチェックを行うことになった。

3 ● 保険診療外の患者（セカンドオピニオンや自費診療）

セカンドオピニオンは、院内がん登録上症例区分8とコードされるため、今回の対象からは自動的に除かれるはずであるが、数例ずつ他のコードがついていることもあった。また自費診療の患者もDPCやレセプトに存在しないということもあった。これらは仕方がないので、対象症例ファイルから各施設で対象症例ファイルから対応する匿名IDを削除してもらうこととした。今回の作業には課題として出てきたが、逆に保険外診療でもきちんと院内がん登録には登録がされているということであり、それは症例捕捉がうまくいっている証拠でもある。

4 ● レセプト上の患者IDが加工されている。

レセプトを代替として提出された施設に関して、正常に匿名化は修了するけれども、レセプトの内容が全て空のファイルとして提出されるという現象が散見された。これはなかなか原因がわか

らなかったが、施設側と何度かやりとりをして、実はレセプト上の患者IDの欄が、IDそのままではなく、「ID（スペース）病棟番号」のように、後にスペースをつけて付加情報をつけている場合が多いということがわかった。特に最近のレセプトコンピュータでそのようなことを行う傾向にあるらしく、システムを入れ替えた付からファイルが全て空になるという現象が起きている。これは、IDに現れるはずのない文字（例えばスペースなど）をソフトで自動判別し、それ以降のIDは無視してマッチングするというやり方で対応することにした。そのような自動判別は理論上可能であるが、現在その作業中で有り、まだ結果は出ていない。

3 これからの方向性と可能性

関連した独立データのリンケージというのは、アイデア自体はどれほど目新しいものではないと考えられるが、実際にやってみるといくつかの現場の課題が浮かび上がってきた。しかし、これらの課題は特に解決不可能な者ではなく、解決したあとには安定してデータを収集する基盤が構築されるものである。本試行については、このような技術を蓄積する重要な意義があったと考えられる。

今後の方向性としてはデータ利用機会の拡大と、収集対象データの拡大の2点が考えられる。データ利用機会の拡大については、現時点では標準実施率を算定して、お互い施設名はマスクした形で比較可能な形で集計し、返却することが主目的となっているが、各種類の患者の診断から治療開始までの日数の算定や、同じ早期患者に対する内視鏡と外科手術の治療選択に関するバリエーションなどが検討できる。また、院内がん登録の初回治療が最終的にどの程度、変更されるのかといった検証も可能である。深く考えずに思いつくだけで解析方法が数多く挙がるので、これが広く研究者に使えるようになれば様々な角度から解析が進みがん対策に資する研究が生まれてくるに違いない。しかし、幅広い活用と個人情報保護のための安全性はトレード・オフの関係であるた

め、その最適な方策を同定する必要がある。また研究者もこのようなデータセットになれている訳ではないので、安全なモデルデータの作成も必要と考えられる。がん研究開発費の研究班ではそのために必要な秘匿化強化法も検討しており、データの安全な活用へ向けた体制整備を検討している。

リンクをする対象データの拡大も今後検討していかなくてはならない。候補としては臓器がん登録や臨床試験のデータなど、情報が詳細で質が高いと期待できるものも同じようなことが可能に違いない。しかし、これらは施設で集められているデータであることが院内がん登録やDPCと共通である。そのため、診療パターンの記述にしても、他院で行われたことが捕捉できないという限界がある。医療連携が推進される中で、この限界は今後大きくなることが予測される。

この限界は、同じレセプトデータでも、健康保険協会や国民健康保険のレセプトからデータが抽出できれば解決する。どこの施設へ行こうとも、保険者へ請求が提出されるため保険者保有のレセプトでは複数医療機関受診の追跡が可能である。もちろん保険者のもつ個人情報を、本来の目的外のために使うわけにはいかないが、今回院内のDPCを匿名化したような技術を応用すれば、自

動的にレセプトデータを保険者から匿名化して抽出することも十分可能である。そのような匿名化により、個人情報保護の問題と、匿名化を行うための作業確保の問題がクリアできれば、データの利用へも道が開けると考えられる。そして、データを十分に解析することができれば、保険者にとっても保健事業の円滑な遂行にもつながり有用である可能性も高い。前例がないことを行うのは障害や懸念も多いだろうが、今や日本は世界にも類を見ない高齢化や医療費の高騰など、既に前例のない時代に突入している。2025年団塊の世代が全て75歳以上となる。そのような時代に備え、新しい発想からそれを裏打ちする地道な技術応用を積み重ねていく活動は必要だろう。

文 献

- 1) National Cancer Intelligence Network. Available from: <http://www.ncin.org.uk/> (Accessed on Dec. 10, 2013)
- 2) Smith M, Saunders R, et al: Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington D.C: National Academy Press, 2012
- 3) American Society of Clinical Oncology. Cancer LinQ. Available from: <http://www.asco.org/quality-guidelines/cancerlinq> (Accessed on Dec.10,2013)